

HEMISPHERIC INVOLVEMENT IN THE PERCEPTION OF SYNTHETIC SYLLABLES,  
NATURAL SYLLABLES, AND "CHIRPS"

BY

MARYLOU PAUSEWANG GELFER

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN  
PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1984

## ACKNOWLEDGMENTS

I would like to thank my chairman, Dr. Harry F. Hollien, for his support of this project, and my committee members for their helpful comments and criticisms. I would also like to express my appreciation to Robert B. Netzer for his programming and extensive assistance with the computer-related aspects of this research. Special thanks go to all my subjects, who endured many hours of electrode placement and re-placement, and endless stimulus tapes. And finally, I would like to thank my husband, John P. Gelfer, for his patience, understanding, and love through a very difficult time.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	ii
ABSTRACT.....	v
CHAPTER I: INTRODUCTION.....	1
The Question of Invariance.....	2
Coarticulation.....	5
Categorical Perception.....	8
Levels of Processing in Speech Perception.....	11
Stimulus Expectation.....	14
Hemispheric Specialization.....	20
A Theory of Speech Perception.....	29
The Problem of Task Variables.....	37
Purpose.....	39
CHAPTER II: METHODS.....	43
Overview.....	43
Methods.....	44
The Principal Components Analysis Procedure.....	70
CHAPTER III: RESULTS.....	75
Preliminary AER Data Analysis.....	75
Analysis One: The Full Data Set.....	77
Analysis Two: Synthetic and Natural Syllables Only.....	115
Analysis of Perceptual Data.....	131
CHAPTER IV: DISCUSSION.....	142
Hemispheric Involvement in Stop Consonant Perception.....	142
Stimulus Expectation.....	148
Hemispheric Involvement in Vowel Discrimination.....	151
Hemispheric Involvement in Stimulus Class Differentiation.....	152
A (Revised) Theory of Speech Perception.....	154
Conclusions.....	160

APPENDIX.....162  
BIBLIOGRAPHY.....166  
BIOGRAPHICAL SKETCH.....175

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

HEMISPHERIC INVOLVEMENT IN THE PERCEPTION OF  
SYNTHETIC SYLLABLES, NATURAL SYLLABLES, AND "CHIRPS"

By

Marylou Pausewang Gelfer

August, 1984

Chairman: Harry F. Hollien, Ph.D.  
Major Department: Speech

The procedures utilized in this research permitted several assumptions underlying a proposed model of speech perception to be tested. It was hypothesized that the voiced stop consonants /b/ and /d/ would be differentiated in the left hemisphere, and that speech and nonspeech tasks involving identical ambiguous stimuli would reveal different perceptual processes and hemispheric asymmetry patterns. Hemispheric involvement in the processing of vowels and acoustic differences between stimulus sets also was examined. Finally, the effect of a task variable--stimulus difficulty/required attention--was assessed. In addressing these issues, average evoked responses (AER's) from the left

and right hemispheres of 12 subjects were collected. The evoking stimuli consisted of the syllables /bi, bæ, bɔ, di, dæ, dɔ/, both synthetic and spoken. In addition, a set of "chirps" (isolated F2-F3 transitions associated with the above syllables) was included. The chirp stimuli were presented twice: once with instructions to discriminate them as /b/ and /d/, and secondly with instructions to discriminate "high" vs. "low" onset frequencies. Subjects indicated whether they heard /b/ or /d/ (or "high" or "low") during the electrocortical recording procedure. The resulting AER's were later analyzed utilizing Principal Components Analysis, Analyses of Variance, and by preplanned and post hoc comparisons. Results revealed an early bilateral differentiation of /b/ and /d/, but inconsistent left hemisphere unilateral processing. Speech vs. nonspeech instructions for identical stimuli elicited dissimilar perceptual strategies; however, differences in hemispheric asymmetry did not reach significance. No evidence for hemispheric asymmetry in vowel discrimination was revealed, although acoustic differences between stimulus sets were discriminated in both bilateral and left hemisphere processes. Finally, stimulus difficulty/required attention appeared to influence patterns of hemispheric involvement. It was concluded that stop consonant perception is mediated primarily through bilateral cortical processes except when discrimination is particularly difficult and that the

perceptual results of this study support the concept of a  
"speech mode" of perception.

## CHAPTER I INTRODUCTION

How is speech perceived? What auditory and neurological mechanisms, what conscious and unconscious strategies must a listener use in order to decode the rapid, complex stream of sound that is speech? Investigations of the speech perceptual processes have revealed that phenomena such as categorical perception and coarticulation appear to be important; and various investigators have hypothesized the existence of a special "speech mode" in which speech stimuli are processed in a different manner than nonspeech auditory stimuli. More recently, the apparent asymmetry of hemispheric function has been discussed in relation to processing the diverse acoustic patterns which comprise speech (Kimura, 1961; Shankweiler and Studdert-Kennedy, 1967; Cutting, 1974; Wood, 1975; Molfese, 1978a, 1978b, 1980a; Molfese and Schmidt, 1983).

Because of the complexity of both the speech signal itself and the human listener, the question "how is speech perceived?" has proven a particularly difficult one to answer. The acoustic structure of the signal must be considered; and within that signal, the specific parameters critical to decoding must be isolated. In addition, the



receptive, perceptual and cognitive structures and processes of the human listener must be taken into account in order to fully describe speech perception. In this research, an attempt will first be made to review the literature pertinent to the cited speech perception problem; and a model suggesting how speech is perceived will be generated. Finally, experiments which extend current findings in this area will be carried out.

#### The Question of Invariance

One of the problems in determining how speech is perceived involves the isolation and identification of consistent acoustic patterns which correspond to particular phonemes. According to Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967), their own initial attempts to isolate speech segments that would be perceived as phonemes were quite unsuccessful, that is except for steady-state portions of vowels and prolonged fricatives. Indeed, the acoustic cues which give rise to the perception of a particular consonant appear to depend greatly on phonemic context.

Early investigations into this issue utilized a speech spectrograph, which displays speech in terms of frequency on the vertical axis over time on the horizontal axis. With such a display, invariant acoustic properties, with the exception of vowels and fricatives, were not observed. On

the contrary, it appeared that in some cases, different acoustic cues were perceived as the same phoneme. For example, the formant transitions cueing /d/ in the syllables /di/ and /du/ were observed to have different frequency compositions and directions, yet both sets of transitions were consistently perceived as /d/ in their appropriate vowel context (Lieberman et al., 1967; Liberman and Studdert-Kennedy, 1978). It was also reported that for the syllable /pu/, the /p/ could be signalled by a rising F2 transition leading into the vowel /u/, while in the syllable /spu/, the /p/ could be signalled by a silent interval of approximately 60 ms between the /s/ noise and the onset of the /u/ (Lieberman and Studdert-Kennedy, 1978). In both cases, two markedly different acoustic cues gave rise to the consistent identification of a particular phoneme. Conversely, a single acoustic cue was observed to signal two different phonemes, depending on context. Cooper, Delattre, Liberman, Borst and Gerstman (1952) describe a study in which a noise burst centered around 1440 Hz was perceived alternately as /p/ or /k/ as a function of the following vowel.

In summary, a simple listing of which acoustic cues correspond to which phonemes did not appear to be an adequate method of addressing the complexity of speech perception. Indeed, some researchers argued that the problem of invariance furnished support for theory of a

"special speech processor" (Lieberman et al., 1967), or a "speech mode" (Mattingly, Liberman, Syrdal and Halwes, 1971), a neural mechanism utilized by a listener during the perception of speech.

More recent evidence based on spectral analysis has begun to show that the problem of invariance may be due more to measurement limitations than actual variations in the acoustic cues for particular phonemes. For example, Stevens and Blumstein (1978) and Blumstein and Stevens (1980) advanced a theory, supported by both pattern-matching of spectra and listener judgements, that the invariant cue for stop consonants is contained within the first "20-odd" ms following the release of the burst. Kewley-Port, Pisoni and Studdert-Kennedy (1983) challenged that viewpoint somewhat, but proposed a very similar theory in which both the static burst information in the first 5 ms following stop release and two additional "dynamic" (or time-varying) cues within the first 40 ms were adequate for listeners to reliably discriminate between stop consonants.

Not all recent research, however, supports theories of invariance in acoustic cues. Howell and Rosen (1983) have shown that the "boundary" between /s/ and /tʃ/, previously hypothesized to be a 40 ms rise time (Gerstman, 1957; Cutting and Rosner, 1974) in fact varies considerably depending on speaking situation in a production task and stimulus range in a perceptual task.

In summary, although invariant cues to individual phonemes may exist within the speech stream, their brief duration and complex nature would appear to make speech perception a difficult task. Early researchers hypothesized an innate "special speech processor" or "speech mode" to account for a listener's ability to extract a particular phoneme regardless of variability in the acoustic signal. Now, it appears that the invariance problem may not be as insoluble as once believed. However, it is possible that the concept of a "speech mode" of perception may still be useful in order to explain the seeming ease with which most listeners are able to decode speech. (See the appendix for a discussion of the nature and possible origin of the hypothesized "speech mode.")

#### Coarticulation

The difficulty in identifying invariant acoustic cues for each phoneme may be due to the dynamic nature of speech and the resulting phenomenon of coarticulation (Ohman, 1966; Daniloff and Moll, 1968; Kuehn and Moll, 1972). In continuous discourse, phonemes are produced in sequences, not as isolated units. As the articulators move to modify the laryngeal spectrum, articulatory positions blend smoothly into one another, and one articulatory position influences another. This influence goes in both directions: the configuration of the oral structure for one phoneme may be

carried forward into the next; or the anticipation of producing a certain phoneme may affect the one which preceded it (Daniloff and Moll, 1968). In any case, the acoustic properties of a particular phoneme depend to a large extent on the surrounding phonemic context--a relationship which results from the overlapping and interrelating movements involved in speech production.

The challenge to perception presented by coarticulation would seem substantial; however, perceptual research has shown that coarticulation actually aids in the rapid perception and processing of speech. Kuehn and Moll (1972) studied listeners' perceptions of portions of spoken CV syllables and found above-chance levels of identification for both phonemes when only the consonant and the initial part of the formant transitions (preceding the vowel formants) were presented. When they contrasted phonemes in terms of manner, voicing and place of production, they found that place of production was the feature most often correctly perceived. Ostreicher and Sharf (1976) examined both forward and backward coarticulatory effects for separated portions of CV, VC and CVC syllables. They found that place of articulation, voicing features and manner features of consonants could be determined from the associated vowel; tongue height and tongue advancement of vowels could be deduced from contiguous consonants; and that subjects were able to determine vowel and consonant

features more correctly from preceding sounds than from following sounds (backward coarticulation). Place of articulation was correctly identified significantly more often than manner in nine of twelve instances. Two of their conclusions were that "coarticulatory effects are perceived and may be used by listeners to help identify adjacent sounds in conversational speech," and that "adjacent phoneme perception involves parallel processing of features" (Ostreicher and Sharf, 1976, pg. 297).

Thus, coarticulation appeared to facilitate speech perception by simultaneously encoding information about several phonemes at any given point in the speech stream. Place information was conveyed particularly well, even when syllables (and thus available cues) were truncated (Ostreicher and Sharf, 1976). Formant transitions appeared to play an important part in perception of articulatory coarticulation, although place information was not coded exclusively through these transitions (Kuehn and Moll, 1972). This knowledge, however, tells us very little about the mechanism which processes speech, except that parallel decoding must be involved. Is a "speech mode" of perception necessary? It can be argued that the acoustic stimuli which comprise speech vary lawfully, depending on coarticulation effects; however, the complexity required to decode the signal in purely acoustic terms is formidable.

### Categorical Perception

Another phenomenon which has been used by researchers to support of a theory of a speech mode is categorical perception. This concept refers to the tendency of a listener to perceive synthetic speech stimuli varied continuously along some dimension as belonging to two or three discrete categories. For example, voice onset time (VOT), the amount of time between the release of a stop consonant and the onset of voicing for the following vowel, serves as a cue for voicing of stop consonants. In English, voiced stop consonants were found to have VOT's of less than 30 ms (depending on the consonant), while voiceless stops generally had VOT's greater than 30 ms (Lisker and Abramson, 1964). According to Liberman et al. (1967), when VOT was varied in 20 ms steps from 0 to 60 ms, English-speaking subjects were generally able to discriminate well between VOT's of 20 and 40 ms, corresponding to the "phoneme boundary" between /b/ and /p/. However, these same listeners were not able to discriminate adequately between VOT's of 0 and 20 ms, or between 40 and 60 ms VOT's, occurring within phoneme categories. Thus, perception of voicing was hypothesized to be categorical: different stimuli belonging to the same category were not discriminated, while stimuli belonging to different categories were discriminated very well along some continuous dimension. Categorical perception has also been

demonstrated for place contrasts through continuous variation of second formant transition (Liberman, Harris, Hoffman and Griffith, 1957). These researchers found that when onset frequency of the second formant transition is varied continuously, listeners do not hear a continuum of step-wise changes, for example, from /b/ to /d/. Instead, as with the VOT-varied stimuli, they perceive the stimuli categorically as belonging to one phoneme class (/b/) or another (/d/), with an abrupt boundary between.

Liberman et al. (1967) contrast the categorical perception that appears to characterize encoded consonants with the continuous perception noted with other acoustic stimuli. They point out that in general, listeners can discriminate about 1200 pitches, although they can identify only about seven. As described above, when the stimuli are consonants, listeners can only discriminate as many as they can identify. These researchers argue that the phenomenon of categorical perception furnishes evidence that speech is perceived and processed in a different manner from other auditory stimuli, and that categorical perception is characteristic of the "speech mode."

It is interesting to note that Liberman et al. (1967) specifically exclude vowels from their categorical perception mode. According to these investigators, vowels are considered long-duration "unencoded" stimuli and can be perceived along a continuum. This hypothesis formed the



basis for a number of challenges to the theories of Liberman and his colleagues, most notably one by Lane (1965). According to Lane's review, if vowels are degraded by being presented in noise, they will also be perceived categorically with steep boundaries similar to those for consonants.

Other investigators have challenged the notion that categorical perception is unique to speech stimuli. Among them are Miller, Wier, Pastore, Kelly and Dooling (1976), who examined a nonspeech noise burst-buzz analog of VOT in consonant-vowel (CV) syllables. These researchers found a sharp boundary and discrimination peaks at category boundaries, similar to the patterns reported for the /pa-ba/ continuum. Pisoni (1977) obtained similar results for two simultaneously presented tones with varying lead or lag times. Cutting, Rosner and Foard (1976) synthesized speech analogs with various rise times, a longer duration rise time which was perceived as a bowed note on a violin ("bow"), and a shorter duration rise time which was perceived as a plucked note on a violin ("pluck"). These investigators also demonstrated that listeners perceived such nonspeech stimuli categorically.

The implication of these categorical perception studies is that categorical perception is not limited to brief-duration consonantal speech sounds, and thus cannot be viewed as evidence for a speech mode of perception.

However, categorical perception remains an important concept in the investigation of speech perception. Brief-duration consonantal stimuli do appear to be perceived categorically, while vowel stimuli (under normal circumstances) do not. Therefore, it would seem that different types, levels or modes of processing do exist, regardless of "auditory" vs. "phonetic" distinctions, and somehow interact during the process of decoding running speech. The nature of these modes of processing, their characteristics and their patterns of interaction have been the subject of a number of investigations.

#### Levels of Processing in Speech Perception

Interference is one type of interaction which can be utilized in the study of processing levels. For example, Day and Wood (1972a) used a reaction time (RT) paradigm to examine the interference patterns of varying vowels and stop consonants. Their results demonstrated that, when two stop consonants occurred in a variable vowel context (/ba, bæ, da, dæ,/), reaction times (RT's) for discrimination were slower than when the same stop consonants were paired with a single vowel (/ba, da/). Conversely, when the target sounds for discrimination were two vowels, reaction times (RT's) were longer when consonant context varied (/ba, bæ, da, dæ,/ ) than when consonant context remained the same (/ba, bæ/). In both cases, interference was mutual. These

findings contrast to Day and Wood (1972b) and Wood (1975), who varied stop consonants and fundamental frequency. In these studies, changing consonant context did not significantly increase RT for fundamental frequency discrimination; however, varying fundamental frequency did significantly increase RT for stop consonant discrimination. In this case, interference was unilateral; i.e., changing the phonetic context (consonants) did not affect nonphonetic discrimination (fundamental frequency), while changing the nonphonetic context did affect phonetic discrimination. Wood (1975) also studied the interference patterns of two nonphonetic aspects of speech: fundamental frequency and intensity. In this context, interference was again mutual, with irrelevant variation of one dimension increasing RT for the target dimension. Collectively, these studies appear to support a hypothesis of two different modes of processing, apparently auditory vs. phonetic, as evidenced by changes in patterns of interference.

Additional research, however, has caused investigators to question the "phonetic" level of processing. Blechner, Day and Cutting (1976) used the "bowed" and "plucked" speech analogs cited above (Cutting, Rosner and Foard, 1976) in a reaction time paradigm. In their study, Blechner et al. (1976) varied rise time and intensity. According to the previous interference studies (Day and Wood, 1972a, 1972b; Wood, 1975), it was expected that rise time and intensity

would show mutual interference patterns, since both are "auditory" (as opposed to "phonetic") stimuli. However, Blechner et al. found that irrelevant variations in rise time did not significantly affect RT for intensity, while irrelevant variations in intensity did affect RT for rise time, a unilateral interference pattern similar to that obtained by Day and Wood (1972b) for consonant vs. fundamental frequency discrimination. From their results, Blechner et al. hypothesized that the "auditory" level of processing most likely consists of several levels; that unilateral interference patterns might be due to ease of discrimination; and that perhaps in previous studies, "the importance of the linguistic-nonlinguistic dimension may have been overrated, and . . . the role of acoustic factors may have been underrated" (Blechner et al., 1976, pg. 264).

Pastore, Ahroon, Puleo, Crimmins, Golowner and Berger (1976) also questioned the notion of a "phonetic" level and, like Blechner et al., studied interference patterns of nonphonetic acoustic tokens. Their stimuli consisted of narrow bandwidth frequency glides (similar to second formant transitions) followed by high and low pitched buzzes (analogous to vowels with different fundamental frequencies). These investigators found unilateral patterns of interference closely resembling those obtained with consonants and fundamental frequency in the Day and Wood

studies. Pastore et al. concluded that their results--as well as those of their predecessors--could be explained in terms of general properties of the auditory system without involving a speech mode of processing.

The two methodologies discussed so far, categorical perception and reaction time, have failed to demonstrate that speech is perceived in a unique and different way from nonspeech stimuli. However, two important points remain to be reviewed. The first is the concept that the unique aspect of speech perception is a function of higher cortical processes; the second is the concept of hemispheric asymmetry and its role in speech perception.

#### Stimulus Expectation

The effects of expectation on comprehension of the more complex linguistic units is a topic that has been discussed frequently, especially by linguists (Chomsky, 1965; Bickerton, 1979; Chu, 1977; Oh and Godden, 1979). The basis of this expectation appears to be semantic and syntactic knowledge of a particular language and cultural conventions. There also is evidence that phonological knowledge of a language establishes certain stimulus expectations and affects auditory perception. For example, Warren and Warren (1970) report research in which certain phonemes in connected discourse were replaced by tones, buzzes or hisses. As a consequence of these operations,

listeners were unable to tell which phonemes had been replaced. They continued to perceive the missing phoneme even after being told where to listen for the substitutions. Although Warren and Warren did not interpret their data in this manner, it is possible that the listeners' knowledge of the structure of English phonology and semantics predisposed them to perceive what they expected to hear.

Day (1970) showed a similar if less consistent phenomenon occurring with dichotic fusion. She took words beginning with consonant clusters such as black, and made two component words, each beginning with one letter of the cluster; for example, black became back and lack. She then presented one component word to a listener's right ear and the other component word to the left. Onset times of the component words were varied such that the back component word was presented with lead times of 100, 75, 50, 25, or 0 ms. Then the lack component word was presented with lead times at the same intervals. However, when the listeners were asked which consonant they heard first (the /b/ or the /l/), two markedly different response patterns emerged. Some subjects consistently reported /b/ as occurring first, even when the lack component word actually led; while others were able to accurately choose which consonant was actually presented first. Interestingly, even when subjects in the former category were told that in some trials the /l/ would precede the /b/, they were still unable to accurately

order the component words. This phenomenon is consistent with the findings of Warren and Warren (1970), where listeners were not able to identify the missing phoneme even when told its general location.

Day's results lead to a number of interesting speculations. First, it appeared that for at least some listeners, knowledge of the phonological rules of English interfered with temporal order judgements. Second, since all listeners did not show this pattern, it might be hypothesized that at least two levels of processing were involved in the task. At one level, phonemes were decoded in the actual order of presentation. At the second level, a reordering of the incoming stimuli occurred for some subjects who appeared to be particularly "language-dependent." This restructuring may have been based on phonological knowledge of acceptable phoneme sequences. The interaction between the second level and the first in the "language-dependent" subjects suggests that higher cortical processes can act in an efferent or down-feeding direction to influence perception. In fact, anatomical evidence for two descending auditory pathways has been found (Harrison and Howe, 1974), and physiological research (Desmedt, 1971; Wiederhold and Kiang, 1970) suggests that cortical involvement can affect sensory processes.

If certain expectations about a stimulus are important influences of perception, the issue of how expectations are

engaged must be considered. On the surface, it would appear probable that acoustic and/or linguistic aspects of the stimulus itself would trigger expectations. However, in the case of ambiguous stimuli, perhaps direct instructions to listeners can be demonstrated to vary mode of perception or level of processing. One example of this technique is demonstrated by Schwab (1981) in a study in which synthesized sinewave analogs of the syllables /bʊ, ʊb, dʊ, ʊd/ (among others) were utilized. These syllables contained appropriate formant frequencies for the vowel and transition onsets, but had bandwidths of 1 Hz. They were not immediately recognizable as speech, according to the author. Schwab instructed half her listeners to discriminate these tokens on the basis of rising or falling frequency, a task presumably requiring an "auditory" mode of perception. The other listeners were told that these tokens were computer-generated speech samples, and were asked to label the tokens, a task requiring a "phonetic" mode. Results of the five experiments contained in Schwab's article consistently indicated that sinewave syllable analogs could be discriminated above chance levels in both modes, and that the discrimination functions for the "auditory" listeners and "phonetic" listeners were different. For example, as additional formants were added, making the signal more speech-like, phonetic discrimination improved while auditory discrimination deteriorated. In addition, there were marked



backward masking effects and frequency masking effects for the auditory group, but not for the phonetic group. Schwab interpreted her results as supporting the concept of a speech mode of information processing. Indeed, her work is particularly important in that she demonstrated that identical stimuli could be perceived in two distinctly different ways, depending on the instructions to the subjects. Expectations about the stimulus (speech vs. nonspeech) appeared sufficient to vary the way an ambiguous signal was processed. However, as hypothesized above, it would seem that the stimulus characteristics themselves are important in engaging or maintaining expectations: the more speech-like the sinewave analogs were, the better they were discriminated in the speech mode.

Other types of ambiguous stimuli also might be used to show how perceptual mode can be influenced. For example, Mattingly, Liberman, Syrdal and Hawles (1971) used a categorical perception paradigm to study perception of second formant transitions, both in isolation and embedded in syllables. Mattingly et al. (1971) found that second formant transitions embedded in syllables were perceived categorically as voiced stop consonants, while those in isolation were not perceived categorically. Further, they reported that the transitions in isolation sounded like clicks or "chirps," and bore no resemblance to speech sounds. In a later study, however, Nusbaum, Schwab and

Sawusch (1983) discussed additional research suggesting that with proper instructions to listeners, chirps might be perceived as speech. Although this was not the focus of their research, they demonstrated that if listeners were told that chirps were parts of phonemes and given practice in identifying them as such, these stimuli could be perceived categorically in a manner similar to intact (synthetic) syllables. Thus, ambiguous auditory stimuli were shifted from being perceived noncategorically in the Mattingly et al. research to being perceived categorically in the Nusbaum et al. study through manipulation of stimulus expectation.

Taken in concert, the results of Schwab (1981), Nusbaum et al. (1983), Warren and Warren (1970) and Day (1970) support a theory of a speech mode of perception and the importance of stimulus expectation in determining how acoustic signals are perceived. Under normal circumstances, this speech mode is probably utilized by a listener when the incoming stimuli have the appropriate frequency and temporal characteristics. However, due to variability in the signal and transmission distortion, the human perceptual mechanism must be capable of processing a degraded signal and still extracting meaning. Thus, stimulus expectation "fills the gaps," and the incoming signal is restructured to conform to some previously learned pattern. This restructuring can occur at many linguistic levels: the perception of a word

or phoneme may be altered to maintain correct syntax or semantic sense (Warren and Warren, 1970); or perception of a various acoustic cues may be altered to maintain a correct phonological sequence (Day, 1970). Thus, the hypothesized "speech mode" may be the result of efferent feedback from the cortical level based on the listeners' expectations about the stimulus.

#### Hemispheric Specialization

In any model of speech perception, it is important to consider the respective roles of the left and right cerebral hemispheres. It has been assumed since the late 1800's that the left hemisphere of the brain is somehow specialized for language (Gevins et al., 1979). Indeed, the pervasiveness of various types of aphasia following injury to or disease of the left hemisphere in right-handed individuals gives credence to this view. But what exactly is the left hemisphere's role in speech perception? The studies cited above appear to indicate the importance of processing levels. If an auditory vs. speech perceptual task could be shown to evoke different patterns of hemispheric involvement, further support would be provided for a theory of different levels or modes of processing.

Unfortunately, "speech" is composed of many acoustically diverse elements, and hemispheric involvement in processing these elements has been difficult to

determine. One popular methodology in the study of hemispheric asymmetry has been dichotic listening. This method was first reported by Broadbent (1954) who found that when two simultaneous competing stimuli were presented binaurally, most right-handed listeners appeared to have a bias for signals coming into the right ear. Kimura (1961) was the first to label this phenomenon as "right ear advantage" or REA. She reasoned that since most of the fibers carrying information from the right ear go to the left side of the brain, this right ear advantage must be indicative of a left hemisphere superiority for processing speech and language. Kimura further hypothesized that these contralateral neural pathways were superior to ipsilateral fibers in conducting sensory information to the auditory cortex.

Right-handed subjects are typically used in lateralization studies, because hemispheric dominance in the left-handed is less predictable (McGlone and Davidson, 1973). However, it should be noted that not all right-handed subjects show REA's for speech stimuli. According to Sidtis (1982), REA is demonstrated by only 70-75% of all right-handed subjects. Further, Sidtis (1982) has hypothesized that only 50% of the dextral population fits Kimura's model of dominant contralateral/secondary ipsilateral pathways. Thus, when applying models of left vs. right hemispheric processing to individuals, it is

important to take into account the range of normal variability. Additionally, when assessing the strength of left or right hemispheric asymmetry as revealed by group trends, it should be remembered that the population mean may include data from subjects with questionable lateralization functions.

Shankweiler and Studdert-Kennedy (1967) investigated REA for syllables consisting of a stop consonant and vowel, and for vowels alone. These researchers found that a right ear advantage existed, but only for CV stimuli; that is, steady-state vowels did not elicit a significant REA.

Cutting (1974) supported and extended the Shankweiler and Studdert-Kennedy findings. While earlier researchers had used only stop consonants in CV stimuli and vowels in isolation, Cutting (1974) included liquids (/r,l/) in addition to stops and vowels. He also examined REA effects related to consonant position, the presence or absence of formant transitions, nonspeech sinewave formant "CV" analogs and inverted or "nonphonetic" formant transitions (labelled as such because they could not have been produced by the human vocal tract). Results revealed that stops had a significantly greater REA than liquids, which in turn had a significantly greater REA than vowels. Both stops and liquids showed an REA, but while final stops retained their REA, final liquids showed an LEA. When results were averaged across conditions, all sounds identifiable as

speech (consonants and vowels) showed some degree of REA. On the other hand, of the sinewave formant CV approximations, those with transitions showed a small, nonsignificant REA, while those without transitions showed a slight LEA. Cutting interpreted these results as supporting a theory that two specific and different mechanisms operate in the left hemisphere, and that both are important to speech perception. The first is involved in processing complex acoustic aspects of the signal, such as the rapid frequency over time changes characteristic of (but not limited to) formant transitions. This mechanism was hypothesized to be "acoustic" in nature rather than "phonetic," since it was activated by nonphonetic (i.e., inverted) transition stimuli, as well as phonetic stimuli. Cutting's hypothesized second mechanism was "phonetic," i.e., a system which responds differentially to speech sounds. He based this second hypothesis on the observation that both CV syllables and vowels with normal speech-like bandwidths evoked an REA, while stimuli that were not perceived as "speech"--sinewave formant CV analogs--did not evoke a significant REA even when transitions were present.

Molfese (1978a) attempted to replicate some of Cutting's results, but employed average evoked responses (AER's) rather than dichotic listening to demonstrate the differential hemispheric responses. Like Cutting, he used stop consonant-vowel syllables with normal (phonetic)

transitions, CV syllables with inverted (nonphonetic) transitions, and sinewave formant stimuli with both phonetic and nonphonetic transitions. Molfese used Principal Components Analysis in analyzing his AER data. This procedure permitted him to identify underlying components of the AER's which might vary with experimental manipulations. Results revealed that both /b-/g/ with phonetic transitions and /b-/g/ with nonphonetic transitions were differentiated in the left hemisphere, but in different ways. Thus, the left hemisphere appeared to be sensitive to normal formant /b-/g/ contrasts, nonphonetic /b-/g/ contrasts, and normal vs. nonphonetic transitions. No such differences were observed in the right hemisphere. It should be noted, however, that the bandwidth variable was not a significant factor in this interaction. That is, in assessing left hemisphere sensitivity to /b-/g/ contrasts, both the responses to normal formant syllables and sinewave formant CV analogs were averaged together. Thus, the results of Molfese (1978a) appeared to support Cutting's proposed left hemisphere mechanism which processed all stimuli containing transitions (although Cutting himself did not demonstrate processing of sinewave formant stimuli in the left hemisphere). Cutting's second hypothesized left hemisphere mechanism which processed only "speech" (normal formant bandwidth speech or speech-like stimuli) was not supported by Molfese's results.

In a 1979 experiment, Molfese and Molfese used stimuli and methods similar to those previously employed by Molfese (1978a), but in this case they used infants of approximately one day old as subjects. Their stimuli consisted of normal bandwidth /ba/ and /ga/ syllables and sinewave formant CV analogs (phonetic vs. nonphonetic transitions were not utilized). The results of this research were similar to those of Molfese (1978a); i.e., both demonstrated a left hemisphere differentiation between the CV syllables /ba/ and /ga/. However, in contrast to the adult subjects, the bandwidth variable was a significant factor in the infants' responses. Results showed that for infants, only /b/ vs. /g/ syllables with normal bandwidth formants were discriminated in the left hemisphere, while sinewave formant syllables were not. Neither were differentiated in the right hemisphere. This study furnished support for Cutting's second proposed left hemisphere mechanism, which processes only stimuli perceived as speech. Molfese and Molfese (1979) attributed this difference in results to possible "maturational" factors, although they did not elaborate on what these factors might be.

One possible explanation for the difference in results between Molfese (1978a) and Molfese and Molfese (1979) is purely acoustic. It is possible that the left hemisphere does indeed process all transitional stimuli, regardless of bandwidth, and does so on the basis of transition direction.



The infants' responses were different from the adults' because presumably they did not have enough experience with acoustic stimuli to respond appropriately to sinewave formant transitions. This explanation, however, is not consistent with Cutting (1974), who did not find left hemisphere processing of sinewave formant stimuli in his adult subjects.

Another possible explanation for the difference in results between Molfese (1978a) and Molfese and Molfese (1979) may relate to the findings of Schwab (1981). As previously reported, Schwab (1981) found that sinewave formant analogs of /ba/ and /ga/ could be perceived either as speech or as nonspeech, depending on instructions to the subjects, as evidenced by differences in discrimination functions. Molfese's (1978a) subjects heard the sinewave formant CV analogs interspersed with normal bandwidth stimuli. Although they were not given specific instructions to label each stimulus as /ba/ or /ga/, it is possible that they did so, thus utilizing the speech mode and differential left hemisphere processing regardless of formant structure. The infant subjects of Molfese and Molfese (1979), however, were not mature enough to employ this strategy, and thus only /ba/ and /ga/ syllables with normal formant structure were differentially processed in the left hemisphere. This interpretation might be reconciled with Cutting (1974) if the different modes of stimulus presentation used by Cutting

(1974) and Molfese (1978a) are taken into account. As noted above, Molfese randomized both his normal bandwidth and sinewave formant stimuli on the same tape; thus subjects heard both types of stimuli in the same trial, and possibly attempted to interpret all stimuli as "speech." The subjects of Cutting, on the other hand, heard all normal bandwidth CV's in one trial, and all sinewave formant CV's in a separate trial, and would therefore have less motivation to try to process all stimuli in the same manner.

Additional studies by Molfese and colleagues (Molfese, 1980a; Molfese and Schmidt, 1983) have generally supported the finding that adult subjects tend to discriminate both normal bandwidth and sinewave formant /b/ and /g/ in the left hemisphere. Molfese (1980a) utilized /b,g/ in varying vowel contexts (/i,æ,o/). Results revealed a significant Hemisphere by Consonant interaction such that /b/ and /g/ (regardless of vowel environment--or formant structure) were differentiated in the left hemisphere but not in the right. Molfese and Schmidt (1983) essentially replicated the Molfese (1980a) preliminary study, reporting similar (though more detailed) results.

Molfese (1980a) and Molfese and Schmidt (1983) were the first AER studies to reveal a consistent left hemisphere response to consonants in varying vowel contexts. This is a significant finding, as the acoustic cues for each consonant are different, depending on the following vowel. However,

in these studies, the effects of several possible confounding factors were not examined. First, transition direction was positive for all /b/ stimuli, regardless of vowel context, and negative for all /g/ stimuli, despite different onset frequencies. Thus, transition direction may have furnished an acoustic cue for consonant identification. Second, subjects' expectations regarding the nature of the ambiguous stimuli were not discussed. In this regard, Molfese and Schmidt (1983) concluded that their results furnished support for a "lateralized mechanism that is sensitive to or extracts relevant linguistic information" (pg. 68). However, it must be assumed that the sinewave formant CV analogs were perceived as speech signals if this conclusion is to be accepted. Finally, it has been demonstrated by Kewley-Port (1982) that formant transitions alone are not sufficient cues in natural speech for accurate stop consonant identification, despite their frequent use in speech perceptual studies. It is possible that perceptual processes identified in the literature could vary significantly as a function of the type of speech stimulus used (synthetic vs. natural).

In summary, the research results reported by Molfese and his colleagues are generally consistent with a theory of a "speech mode" of perception. This speech mode is characterized by left hemisphere differentiation of stimuli containing transitions and which subjects perceive as

"linguistic," regardless of acoustic differences in cues related to varied vowel contexts. However, the results of Molfese (1978b) are contradictory to a theory that the speech mode is a left hemisphere function.

Molfese (1978b) has suggested that one of the acoustic cues for voicing of stop consonants, voice onset time (VOT), appears to be processed in the right hemisphere. That is, in a typical categorical perception paradigm, a differential response to between-category VOT changes (20 and 40 ms) was only found in the right hemisphere. To be specific, the right hemisphere response correlated with listeners' perception of /b/ vs. /p/, while the left hemisphere did not. However, differential left hemisphere responses to the endpoints of the continuum (0 and 60 ms) were observed; and a second response showed discrimination of the endpoints (0 and 60 ms) from the midpoints (20 and 40 ms) of the continuum. Similar results were obtained by Molfese (1980b) when nonspeech tonal stimuli with varying relative onset times were utilized. Thus, the idea of a simple correlation between a speech mode of perception and left hemisphere activity appears to be inadequate to describe the actual complexity of speech perception.

#### A Theory of Speech Perception

While a theory explaining speech perception is desirable, it should be one which includes hemispheric

asymmetry data and the concept of stimulus expectation. Such a model is presented in Figure 1-1.

#### The Feature Level

As may be seen, the model specifies that both the left and right cerebral hemispheres are active in the primary processing of acoustic stimuli. Complex, rapidly-changing frequency over time information (including formant transitions) is analyzed in the left hemisphere. Left hemisphere involvement in the processing phonetic and nonphonetic transitions has been demonstrated by Cutting (1974), and Molfese (1978a), although evidence for similar asymmetrical differentiation of sinewave formant stimuli is less clear (Molfese, 1980a; Molfese and Molfese, 1979; Molfese and Schmidt, 1983). At the same level, further analysis of the spectral and temporal characteristics of the acoustic signal may take place in the right hemisphere. These perceptual processes in both the left and right hemispheres could be considered the "feature level" of speech perception, because decisions as to place, manner and voicing are made at this point, in accordance with feedback from higher cortical processes. If the stimuli were not speech-like, a different set of expectations would be utilized by the listener, and another model would be necessary.

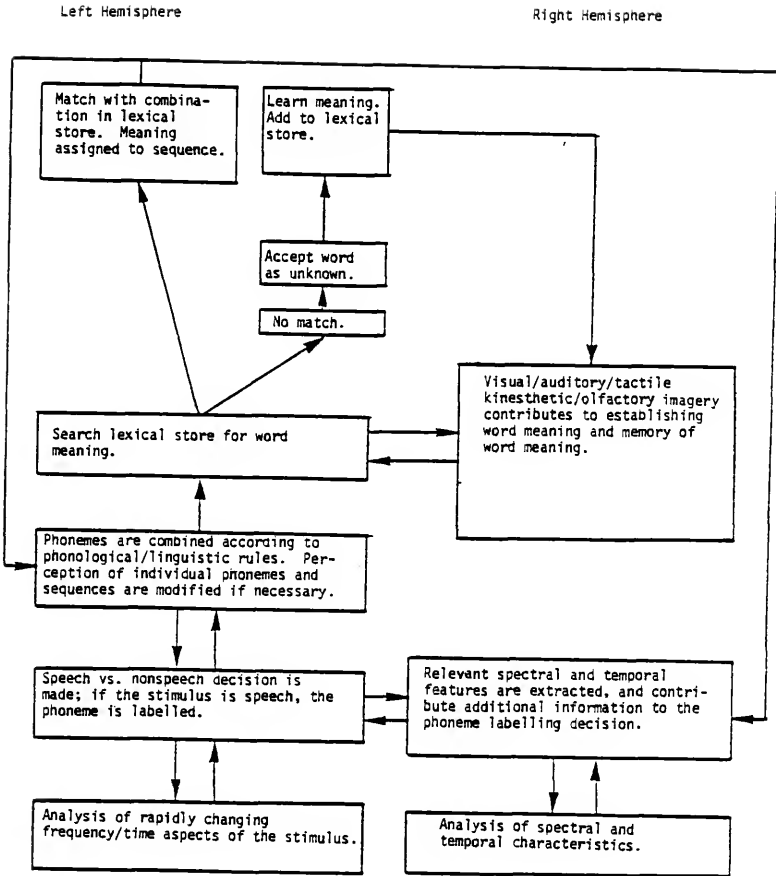


Figure 1-1. A model of speech perception. See text for discussion.

### The Phoneme Level

At the next level, the feature information is combined in the left hemisphere, and it is determined if the stimulus is speech or some nonspeech signal. Again, a basis for this mechanism can be seen in Molfese (1978a), where both phonetic and nonphonetic /b/ vs. /g/ were discriminated differently in the left hemisphere. Stimuli presumably would be classified as speech or nonspeech on the basis of 1) frequency, bandwidth and other characteristics of the signal (Schwab, 1981), 2) the presence of acoustic cues appropriate to a particular place or manner of articulation and voicing, and 3) the temporal characteristics of the signal. However, even in the absence of clear cues for speech, feedback from higher cortical centers may override the inadequate acoustic cues, and the signal may be perceived as "speech" (Schwab, 1981; Nusbaum et al., 1983). Simultaneous with the speech-nonspeech decision is one regarding the identity of the phoneme. Again, stimulus expectation, now based on linguistic knowledge, may override actual acoustic cues. Expectations at this level may also include some knowledge of how speech is produced; thus, if an articulatory referent exists (Liberman et al., 1967), the left hemisphere phoneme level is the processing stage where such cross-correlations would take place in this model.

In the right hemisphere, initial signal processing at the feature level involves the spectral and temporal aspects of a signal, as mentioned above. At the phoneme level, the relevant features extracted from such an analysis are transmitted to the left hemisphere. Thus, right hemisphere input contributes to the speech vs. nonspeech decision and phoneme labelling. There can also be feedback from the left hemisphere to the right hemisphere at this level if the feature information does not conform to expectations or if it results in an ambiguous speech/nonspeech decision or phoneme label. There is constant interaction between the hemispheres at this level. These processes take place below the level of consciousness.

#### The Word Level

At the word level of processing, phonemes are combined and sequenced in the left hemisphere according to phonological rules. Marked individual differences can be seen at this level, with some subjects' perception highly dependent on phonological knowledge while others' perception is not (Day, 1970). A number of auditory "illusions" may also occur (Warren and Warren, 1970) when the actual acoustic cues for a particular phoneme are omitted or distorted. Again, there is constant feedback between the phoneme and word levels of processing. If the incoming phoneme sequence violates phonological or semantic rules, the sequence of phonemes may be altered in the left



hemisphere, or the questionable sound may be shunted back to the phoneme level to be relabelled. At this point, there is also interaction between the hemispheres. Feedback from the left hemisphere to the right can influence further spectral and temporal analysis, while feedback from the right hemisphere to the left can influence phoneme sequencing. This level is at the borderline of consciousness; thus, some listeners may be aware of modifications while others are not (Day, 1970).

#### The Concept Level

At this juncture, a particular meaning must be associated with the sequence of phonemes--a left hemisphere function. The "lexical store," or long-term word memory, is searched for similar phoneme sequences. If such a sequence is found, the same meaning is assigned to the incoming stimulus item. If a similar sequence is not found, several alternate steps may ensue. The listener may accept the word as an unknown, and process no further. Or, the word may be stored while the meaning is gradually learned, thus adding to the lexical store. Alternatively, the listener may try to find the best possible "match" already existing in the lexical store, and accept the sequence as a known but distorted word. The latter occurs when attempting to understand a speaker with a speech problem or foreign accent. At this point of processing, modifications are easily accessible by the conscious mind, although individual

variables such as intelligence, training and motivation determine the extent of conscious involvement.

In the right hemisphere at the conceptual level, sensory imagery contributes to establishing word meaning and memory. The listener associates information from the visual, auditory, tactile, kinesthetic and olfactory modalities with phoneme sequences in order to fully comprehend the meaning of a particular word. Interaction between the left and right hemispheres at this level serves as the link between external language and internal representations.

As the final step in a feedback loop, the acoustic characteristics and features of the word are channeled to lower levels in both hemispheres. In the left hemisphere, this information influences the manner in which future phoneme input will be sequenced and modified. In the right hemisphere, this feedback will affect the phoneme level and future extraction of salient spectral and temporal features.

This model can be considered the "speech mode" of perception. Although previous investigators have demonstrated that adequately complex nonspeech stimuli can evoke similar perceptual processes, only with speech stimuli do expectations regarding syntactic, semantic, phonological and possibly physical constraints feed downward through the left hemisphere to affect perception at a basic level.

Most of the literature cited in this paper supports this model, particularly in terms of the importance of stimulus expectation and efferent feedback. Day (1970) has shown that a specific sequence of phonemes can be unconsciously reordered by a listener in order to conform to English phonological rules. Warren and Warren's (1970) results indicate that missing phonemes can be perceived as being present, presumably based on listeners' expectations. Finally, Nusbaum et al. (1983) and Schwab (1981) have demonstrated that identical stimuli can be perceived in distinctly different ways based on instructions to the listeners for engaging different sets of expectations. On the other hand, Molfese's (1987b) results are not completely consistent with the proposed model. According to this model, /ba/ and /pa/ should have elicited some differential processing in the left hemisphere, since they are presumably processed with reference to stimulus expectation. No such left hemisphere involvement at the phoneme boundary was discovered. However, it may be the case that since place of articulation (and therefore second formant transitions) was the same for both consonants, the left hemisphere did not differentially process these syllables; while the temporal processing in the right hemisphere based on learning of the appropriate VOT's in English did discriminate between the two in this particular study. In contrast, research in aphasia (Gandour and Dardarananda, 1982) has revealed that

patients with left hemisphere lesions were significantly impaired in VOT perception. This would tend to confirm the importance of the left hemisphere phonetic labelling features, and in perceiving a signal as speech. Finally, as noted previously, the results of Molfese (1978a, 1980a) and Molfese and Schmidt (1983) are not completely compatible with the hypotheses advanced in this model unless one accepts the premise that their sinewave formant CV analogs were perceived by listeners as "speech." Research which includes manipulation of subjects' expectations regarding the "speech" or "nonspeech" nature of identical ambiguous stimuli is needed to clarify this issue.

#### The Problem of Task Variables

Much of the research cited in support of the model presented above involves the AER methodology and the work of Molfese and colleagues (Molfese, 1978a; 1980a; Molfese and Molfese, 1979; Molfese and Schmidt, 1983). Their research generally included both normal syllables and nonspeech CV analogs; and in their analysis, responses from both sets of stimuli were averaged together. This research design, however, does not take into account two important variables: first, one set of stimuli is less familiar than the other, more difficult to discriminate, and presumably requires a greater degree of attention from subjects; and second, some

of the subjects' perceptual judgements of the ambiguous stimuli will be incorrect.

Regarding the differences in stimulus difficulty/required attention, the results of numerous AER studies suggest that as difficulty in discriminating among stimuli increases, AER latencies become longer (Ritter, Simson and Vaughn, 1972) and amplitude increases (Poon, Thompson and Marsh, 1976). Other studies have shown that as the amount of attention required by a task increases, so do AER amplitudes (Eason, Harter and White, 1969; Harter and Salmon, 1972). Further, dichotic listening studies have shown that increasing task difficulty results in larger hemispheric differences. For example, when listeners were asked to identify vowels in noise (Weiss and House, 1973) and vowels of brief duration (Godfrey, 1974), a tendency toward right ear advantage increased. Further, Kasischke (1979) demonstrated that increasing the complexity of tonal stimuli resulted in asymmetric left hemispheric involvement. Thus, it is possible that the left hemisphere /b-/g/ discrimination found in the Molfese research may be dependent upon the inclusion of ambiguous stimuli in the research design, and does not reflect normal speech perception.

The second confounding variable mentioned above, incorrect perceptual judgements, is also potentially serious. If one assumes that electrocortical activity

reflects a cognitive process or series of processes, an incorrect perceptual judgement should result in a slightly different waveshape than a correct perceptual judgement. Thus, it would appear important to include only correct perceptual judgements when averaging trials to obtain AER's.

In summary, it is possible that uncontrolled task variables affected the results obtained in previous AER studies. A research design which includes stimulus difficulty/required attention and accuracy of judgement as independent variables would appear to be necessary in order to separate hemispheric response to stimulus characteristics from hemispheric response to task variables.

#### Purpose

The purpose of this study is to test several aspects of the theory of a speech mode of perception presented above. According to this theory, perception of stop consonants should result in a cognitive process specific to the left hemisphere. Further, when ambiguous stimuli are utilized, subjects who have been instructed to perceive these tokens as "speech" should demonstrate a similar left hemisphere differentiation. When subjects are instructed to process the same stimuli in a nonspeech manner, a different pattern of hemispheric involvement is predicted.

In addition to testing these hypotheses, a number of more general questions related to inter- and

intra-hemispheric processing of the various classes of stimuli will be explored. They are: 1) Are there bilateral processes which differentiate /b/ from /d/, regardless of vowel context? 2) Do AER's from the left and right hemispheres differ significantly, regardless of consonant, vowel or trial? 3) Are there bilateral processes which discriminate vowels, regardless of consonant context or trial? 4) Are there bilateral processes which discriminate between trials (natural syllable trial, synthetic syllable trial, chirps with speech instructions, chirps with nonspeech instructions)? 5) Is /b/ differentiated from /d/ in the left hemisphere regardless of trial? 6) Do trials appear to be discriminated in one hemisphere or the other? 7) Is there any evidence for hemispheric asymmetry in the perception of vowels?

Finally, an additional purpose of this study is to explore the effect of two task variables (stimulus difficulty/required attention and correct judgements) on the obtained pattern of cortical responses. In order to assess the importance of these task variables, the data obtained in this study in response to synthetic syllables and natural syllables will be analyzed separately from the chirp data. With the exclusion of the chirp trials and incorrect syllable perceptions, only stimulus presentations in which subjects correctly judged /b/ or /d/ will be included when calculating individual AER's.

The primary hypothesis to be tested in this second analysis is that /b/ and /d/ will be significantly different in the left hemisphere, but not the right, for both the natural and synthetic syllables. Such a finding would support a theory of stimulus expectation, and reject a hypothesis that stimulus difficulty/attentional variables caused the left hemisphere differences noted in the previous research.

As in the first analysis, a number of secondary questions will also be considered. These are: 1) Is there a similar pattern of hemispheric involvement for /b/ vs. /d/ discrimination for both synthetic syllables and natural syllables? 2) Are there bilateral processes which differentiate /b/ from /d/, regardless of vowel context or trial? 3) Are there bilateral processes which differentiate synthetic and natural syllables? 4) Are there left or right hemispheric processes which differentiate between the two types of syllables? 5) Are there bilateral processes which differentiate vowels regardless of consonant context? 6) Is there any evidence of hemispheric asymmetry in the perception of vowels? 7) Are there significant differences between the AER's from the left and right hemispheres regardless of consonant, vowel or trial?

Finally, this study will examine subjects' perceptual responses in the two chirp trials (speech vs. frequency instructions). Accuracy of response between the two trials



will be compared, both as a main effect and as a function of the order in which the instructions were presented to subjects. Error patterns between the two trials will also be compared as a main effect and as a function of order of instructions. Finally, subjects' perceptions of their strategies for discriminating between the two classes of stimuli in the "speech" trial and the "nonspeech" trial will be informally compared.

## CHAPTER II METHODS

### Overview

The purpose of this study was to investigate hemispheric involvement during the perception of phonemes--specifically, stop consonants--and to evaluate a theory of a "speech mode" of perception. Cortical responses were collected from twelve subjects in response to both synthetic and spoken (natural) /bi, bæ, bɔ, di, dæ, dɔ/, and to isolated F2-F3 transitions ("chirps"). In one chirp trial, subjects were instructed to label the stimuli as beginning with /b/ or /d/; and in a second, they were instructed to listen for "high" vs. "low" onset frequencies. Subjects' cortical responses from the left and right hemispheres were digitized, averaged and normalized on a PDP 11/23 computer. The resulting average evoked responses (AER's) were later subjected to off-line Principal Component Analysis. The resulting factor scores were used as dependent variables in a number of Analyses of Variance in order to determine if any changes in AER could be related systematically and significantly to the independent variables (hemispheres, consonants, vowels or trials).

MethodsStimuli

Stimuli of three types were utilized in this research. They included synthetic syllables, natural syllables and "chirps," or isolated F2-F3 transitions.

The synthetic stimuli were six CV syllables, /bi, bə, bɔ, di, dæ, dɔ/; each consisted of a 50 ms transition followed by a 300 ms steady-state segment. These vowel and transition durations parallel those reported by Cutting (1974), Molfese (1978a, 1980a) and Molfese and Schmidt (1983). They were used in this study in order to facilitate cross-research comparisons. Specific onset values for each transition and the associated steady-state formant are given in Table 2-1. Transition onset frequencies were taken from data presented by Kewley-Port (1982) and Klatt (1980), and modified as necessary during synthesis in order to achieve optimal discriminability. Vowel formant frequencies for F1 through F3 were taken from Peterson and Barney's (1952) data. It will be noted that F4 and F5 are constant across the entire syllable duration, and are the same for each vowel. The upper formants were included in order to make the synthetic syllables sound more natural. For all syllables, bandwidth of F1 was 60 Hz, for F2, 90 Hz, and for F3 through F5, 120 Hz (Cutting, 1974; Molfese, 1978a, 1980a; Molfese and Schmidt, 1983). Each syllable had an

Table 2-1. Onset and steady-state frequencies for synthetic syllables.

Syllable	Formant	Onset Frequency (Hz)	Steady-state (Hz)
/bi/	F1	200	310
	F2	1100	2020
	F3	2150	2960
	F4	3300	3300
	F5	3750	3750
/bæ/	F1	200	620
	F2	1100	1660
	F3	2150	2430
	F4	3300	3300
	F5	3750	3750
/bo/	F1	200	600
	F2	900	990
	F3	1900	2570
	F4	3300	3300
	F5	3750	3750
/di/	F1	200	310
	F2	1800	2020
	F3	2960	2960
	F4	3300	3300
	F5	3750	3750
/dæ/	F1	200	620
	F2	1600	1660
	F3	2700	2430
	F4	3300	3300
	F5	3750	3750
/do/	F1	200	600
	F2	1600	990
	F3	2700	2570
	F4	3300	3300
	F5	3750	3750

associated fundamental frequency of 130 Hz (Peterson and Barney, 1952), and a rise time of 30 ms.

All synthetic syllables were produced by a Klatt software synthesizer (Klatt, 1980) implemented by a Data General IV computer, digital to analog converter and low-pass filter with a cutoff frequency of 5000 Hz. Stimulus parameters were entered using a Hewlett-Packard 2648A Graphics Terminal. All syllables were recorded on one channel of a TEAC 6120 dual channel tape recorder. See Figure 2-1 for the equipment configuration.

The natural syllable stimuli were produced by a male speaker with clearly identifiable vowel formants and the ability to modify fundamental frequency upon request. During production of the stimuli, the speaker was seated in a double-walled Industrial Acoustics Company (IAC) booth. Stimuli were recorded using a B&K 5065 half-inch condenser microphone and a B&K 37A preamplifier, coupled with a Revox B-77 tape recorder. First, the speaker produced each syllable five times. Each of the recorded syllables was then examined on a Voiceprint Model 700 t-f-a spectrograph for vowel and transition durations, and clarity and stability of formant structure. At this juncture, the best two or three examples of each syllable were modified by eliminating prevoicing of the consonant and by reducing vowel duration to conform as closely as possible to the synthetic stimuli (50 ms transitions, 300 ms vowel

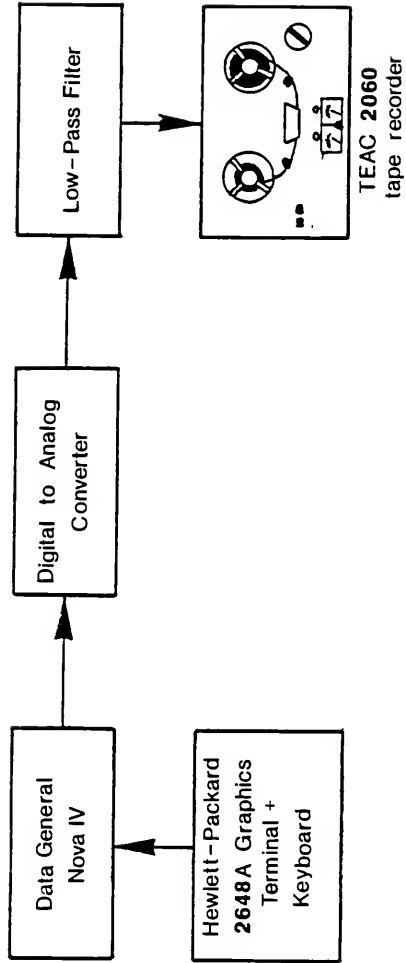


Figure 2-1. Equipment configuration for generation of synthetic speech.

durations). Mean transition duration of the selected syllables was calculated to be 48 ms (range: 24-72 ms). Mean vowel duration was found to be 292 ms (range: 262-304 ms). Finally, the rise time of each syllable was calculated from the output of a Honeywell 1508 A Visicorder, and the exemplars with a rise time most closely approximating 30 ms were selected for inclusion as natural syllables stimuli. Mean rise time of the selected syllables was 38.1 ms (range: 31.2-41.6 ms). Actual onset (transition) and steady-state (vowel) frequencies for the first three formants of the natural syllables stimuli are provided in Table 2-2.

The "chirps," or isolated F2-F3 formant transitions, were utilized as ambiguous stimuli in this study. They were selected because such brief-duration signals do not sound at all like "speech" to naive listeners, and most probably would not be perceived as speech without special instructions. Thus, stimulus expectation could be controlled through instructions to the subject.

The chirp stimuli were taken from their respective synthesized complete syllables. When the program for each syllable was run, only the first 50 ms (the transition portion) was activated. A digital filter developed by J. J. Yea at the University of Florida was utilized in order to eliminate fundamental frequency, F1 and F4-5. The isolated F2-F3 transitions, or "chirps," for each of the six syllables were recorded in the manner described above.

Table 2-2. Onset and steady-state frequencies for natural syllables.

Syllable	Formant	Onset Frequency (Hz)	Steady-state (Hz)
/bi/	F1	250	313
	F2	1800	2000
	F3	2375	2850
/bæ/	F1	625	750
	F2	1500	1625
	F3	2313	2450
/bo/	F1	500	625
	F2	938	1000
	F3	2225	2225
/di/	F1	188	313
	F2	1937	2125
	F3	2765	3000
/dæ/	F1	380	750
	F2	1700	1375
	F3	2650	2375
/do/	F1	385	610
	F2	1500	1000
	F3	2480	2225



### Tape Construction

Three stimulus tapes were constructed for presentation during the experimental sessions. One contained the synthetic syllables, one the natural syllables and a third the chirps (F2-F3 transitions in isolation). Each of the six syllable (or chirp) stimuli was repeated 20 times in random order on each tape, for a total of 120 stimuli per tape. A random numbers table was utilized in establishing the stimulus sequence. The original order was maintained on all three tapes due to program limitations.

The specific recording procedures for each tape were as follows: the 120 stimuli were recorded in the specified sequence on both channels of an Akai GX-77 tape recorder from the master tape played on a Revox B-77 tape recorder. Inter-stimulus intervals were varied from two to nine seconds in order to avoid eliciting a time-locked cortical expectancy response from the subjects. Maximum amplitude of each syllable or chirp was monitored on the VU meter of the Akai tape recorder, and adjusted prior to recording so that all stimuli peaked at 0 VU.

### Subjects

Subjects were 12 young adults--six males and six females--aged 23-33 years. The mean age for male subjects was 28.2 years with a range of 23.3 to 32.8 years; and mean age for females was 27.8 with a range of 23.4 to 30.7 years. All subjects were majoring/employed in the fields of

experimental phonetics or speech pathology, and participated in this study at the request of the experimenter.

In the first selection protocol, subjects were required to demonstrate pure tone thresholds of better than 20 dB at 0.5, 1.0, 2.0, 4.0, and 8.0 kHz, with a mean between-ear threshold difference of less than 5 dB. In addition, any potential subject with a 10 dB or greater between-ear threshold difference at any single frequency was rejected. These criteria were included in order to eliminate potential asymmetric hemispheric effects due to failure to control for differences in peripheral sensation level. Subjects also were selected on the basis of a strong right-hand preference as measured by the Edinburgh Inventory of Handedness, or EIH (Oldfield, 1971). This second selection protocol was included because previous research had indicated that there might be some interaction between hand preference and cortical responses to syllables (Molfese, 1978a). The study of such an interaction is undoubtedly important for refining and generalizing theories of speech perception. However, an investigation of individual differences in perceptual asymmetries as a function of handedness is not the focus of this research. Subjects were limited to those demonstrating a strong right hand preference with the assumption that such an effect reflects a dominant left hemisphere. After a model of hemispheric involvement in speech perception has been established for this population, modifications of the

model may be added through further research involving sinistrals, ambidextrals and other groups of questionable cerebral dominance. In the present study, the average laterality quotient on the EIH was 92.8 (range: 83-100), with a mean decile of 8.21 (range: 6-10). These scores suggest a strong right-hand preference in the subjects utilized in the experiment. As a final step in the selection process, a screening test of the synthetic stimuli was presented. A tape was played, containing ten randomized samples of each syllable; and subjects indicated which consonant they heard at the beginning of each stimulus item. This protocol was included in order to insure that these stimuli were perceived correctly. A score of 95% or better on the 60-item screening test was required in order for volunteer subjects to be included in the experiment. Subjects were allowed up to three attempts to pass the test. On the final trial, mean percent correct was 98.5% with a range of 97-100%.

#### Procedure

The experimental procedure included a training protocol prior to presentation of the syllable stimuli, determination of electrode locations and placement on the subjects' heads, actual electrocortical recording of the subjects' responses to the syllable stimuli, a second training protocol prior to presentation of the chirp stimuli, and electrocortical

recording of subjects' responses to the chirps. This procedure took approximately five hours for each subject.

Training prior to presentation of syllable stimuli.

Subjects were seated individually in a double-walled IAC booth and familiarized with the testing environment. At this juncture, they were instructed in the response procedure. As has been discussed, it was considered important to monitor the accuracy of subjects' perceptual judgements; thus, subjects were required to make an overt response to each stimulus presented.

The response procedure involved use of a Wollensak 4055 battery-powered tape recorder coupled to a microphone. Subjects were instructed to hold this microphone in a comfortable position such that the index finger of one hand rested on it while the fourth finger of either hand did not make contact with the microphone in any way. When subjects perceived the first phoneme of the syllable as /b/, they were instructed to gently raise and lower the index fingers of both hands. Subjects were further instructed to respond to syllables beginning with /d/ by gently raising and lowering the fourth fingers of both hands. A bilateral motoric response was judged necessary in order to eliminate potential hemispheric asymmetry associated with a unilateral response. This procedure resulted in a sound being recorded on the Wollensak in response to stimuli perceived as /b/, and no sound in response to stimuli perceived as /d/. This

code was later utilized by the experimenter in determining correct and incorrect perceptual responses.

At this time, the screening test for synthetic syllables was administered. If potential subjects exhibited more than three errors in the 60 trials, they were permitted to listen to the training tape a second time, then took the test again. If they failed the screening test a second time, they could choose to terminate their participation in the study or return a third time for a last attempt.

A 60-item screening procedure for the natural stimuli was also administered in order to familiarize subjects with these experimental syllables. As in the synthetic syllable screening procedure, if any subject had been unable to achieve a score of 95%, they would have been eliminated from further participation. However, these stimuli were not difficult to discriminate, and no subject exhibited any difficulty whatsoever with this set of protocols.

Electrode placement. The active electrode sites chosen for this study were T3 and T4 as described in Jasper's (1958) "10-20 Electrode System." These locations were chosen because they (theoretically) overlie the left and right posterior superior temporal gyrii, areas of the brain associated with primary auditory reception (Penfield and Roberts, 1959). Additionally, recent AER research has shown that right and left hemisphere differences can be observed at those locations (Molfese, 1978a; Molfese, 1980a;

Molfese and Schmidt, 1983; Wood, 1975). Finally, use of standardized electrode placements within the 10-20 System was judged desirable in order to facilitate interlaboratory comparisons of data.

In the 10-20 System, recording sites are located either 10% or 20% of the distance between several standard reference points for measurement. These standard points are the nasion, or bridge of the nose; the inion, or occipital protuberance; the left and right aural clefts (A1 and A2 respectively); and CZ, the intersection of a line drawn from the nasion to the inion with another from A1 to A2. The T3 location as described by Jasper (1958) is 10% of the distance from A1 to A2 as measured upward along a line from A1 to CZ. The T4 location was measured the same way except from A2 to CZ (see Figure 2-2). These locations are designated "T" because they are assumed to overlie the temporal lobe (anatomical studies presented in Jasper, 1958, support this assumption). The "T3" location denotes the left hemisphere, as all odd numbers are on the left side of the head, while the "T4" location denotes the corresponding point on the right hemisphere.

The active electrodes (T3 and T4) were referenced to contralateral earlobes. These inactive sites were selected because there is little muscle tissue in that area to generate EMG artifacts, and they are less subject to picking up temporal lobe activity than mastoid sites (Goff, 1974).

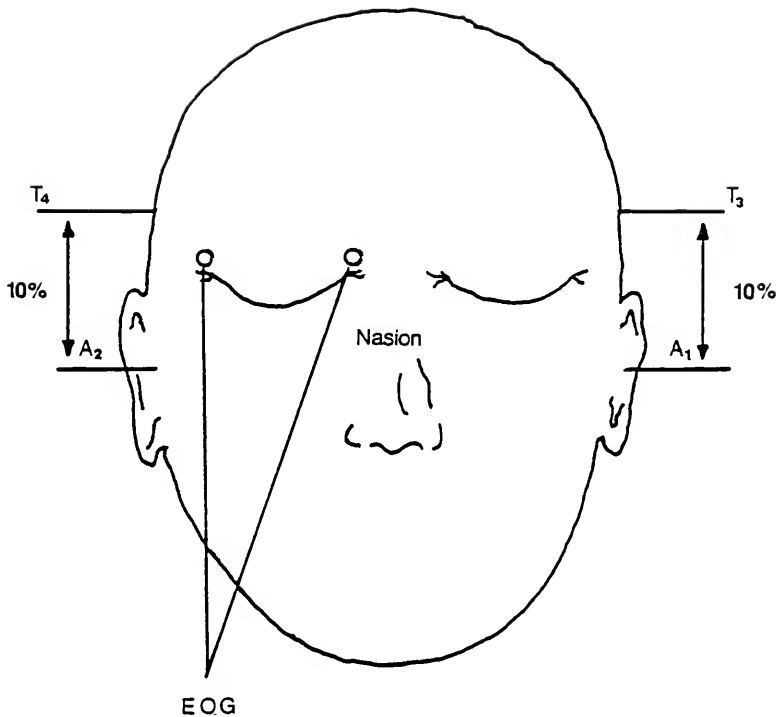


Figure 2-2. Electrode configuration. Electrocortical responses were recorded at  $T_3$  and  $T_4$ . The EOG electrodes recorded eyeblinks and facial movements for input to an artifact rejection channel.

Additionally one electrode was placed above the inside corner of the subject's right eye while another was placed at the lateral superior aspect of the right orbital ridge. These electrodes were used to record extraocular eye movements (extraocculogram, or EOG) and blinks, for an artifact rejection channel. Finally, one electrode was placed on the left mastoid process to serve as a grounding electrode.

Once the seven electrode sites were located and marked, the skin at each was thoroughly cleaned with a cotton swab dipped in alcohol. This type of cleansing is necessary to remove any skin oils or dead epithelial cells which reduce electrical conductivity. Grass E6SH chlorided silver electrodes were then filled with paste (Grass EC-2) and attached to the subject's head with surgical tape. This type of electrode has been recommended for recording "slow" electroencephalographic (EEG) waves because of the resistance of this combined substance (silver-silver chloride) to polarization (Goff, 1974).

Next, resistance in kOhms was measured between 1) T3 and the right earlobe, 2) T4 and the left earlobe, and 3) the two EOG (facial) electrodes, both prior to the recording session and at its conclusion. For the active electrodes, resistances were as follows: for T3, average initial resistance was 3.82 kOhms (range: 1.1-8.9 kOhms); for T4, average initial resistance was 4.03 kOhms (range: 1.5-6.4



kOhms). Final resistances measured at the end of each recording session averaged 4.39 kOhms for T3 (range: 1.3-11.1 kOhms) and 4.59 kOhms for T4 (range: 1.7-7.8 kOhms).

Training and presentation procedures for syllable stimuli. Subjects reclined on a bed in a comfortable position in a double-walled, electrically shielded IAC booth during the electrocortical recording protocol. They were instructed to keep their eyes closed, jaws relaxed and move as little as possible during the stimulus presentation in order to minimize movement artifacts. Subjects were provided the microphone of a small battery-powered tape recorder and reminded to raise both index fingers if the stimulus item initiated with a B and both fourth fingers if the stimulus item began with a D, as they had done during the training procedures. Presentations were counterbalanced; that is, the synthetic stimuli were presented first to half the subjects, and the natural syllables first for the other half. All subjects were permitted a short break following the presentation of the first set of syllable stimuli, and electrode resistances were checked. This procedure was carried out in order to insure that the electrodes were still properly attached and in good contact with the scalp. Subjects then returned to the booth for the second set of syllable stimuli.

Training and presentation procedures for the chirp stimuli. At the conclusion of the second syllable trial, subjects were again given a short break and the electrode resistances tested. At this juncture, subjects were presented stimuli from a second training tape in order to familiarize them with the chirp stimuli. Half the subjects were first instructed that the chirps were frequency glides, and they were to discriminate high vs. low onset frequencies; while the other half of the subjects were first instructed that the chirps were parts of syllables and they were to discriminate /b/ from /d/. It should be emphasized that the stimuli in both trials were exactly the same; only the instructions varied. All subjects were given both instruction conditions, with order of presentation balanced across subjects. After the first chirp trial, subjects were given a short break, electrode resistances were checked, and the second training tape was played. They then returned to the booth for the last trial. At the end of the session, electrode resistances were measured one final time.

Electrocortical recording. The procedures followed during electrocortical recording and stimulus presentation were as follows: stimuli from the right channel of an Akai GX-77 tape recorder were played through a Kenwood KA-7100 amplifier outside the booth to an ADS 810 speaker located inside the booth, at an intensity level of 62 dB re: .0002

d/cm<sup>2</sup> at the subject's ear. The speaker was positioned approximately 78 inches directly in front of the subject. The syllables or chirps on the left channel of the stimulus tape were input directly to a Schmidt trigger, which produced a 4 V pulse at the onset of each syllable. This pulse was utilized to synchronize stimulus onsets during the averaging procedure.

During the cortical site recordings, two Grass 7P122A Low Level DC Amplifiers switched to AC settings were used. Bandpass was flat (half amplitude) from .04 Hz to 60 Hz. This bandpass setting insured that frequencies from .3 Hz to 35 Hz would be amplified at 100% of maximum gain. Such a range was desired in order to maximally amplify all frequencies which might be associated with syllable discrimination, while attenuating the very slow (DC) potentials associated with the contingent negative variation and the very high frequencies associated with electrical interference. System gain was set at 28k, in order to amplify the raw EEG wave to +/- 1.25 V, the optimal range for input to the A/D converter.

For the EOG (facial) electrodes, a Grass 7P3B AC Preamplifier coupled with a Grass 7DAF DC Driver Amplifier was used. Bandpass was flat (half amplitude) from .3 Hz to 75 Hz, with gain set at 11k. Because the data from this channel served only for artifact rejection purposes, bandpass and gain settings were less crucial. Gain was

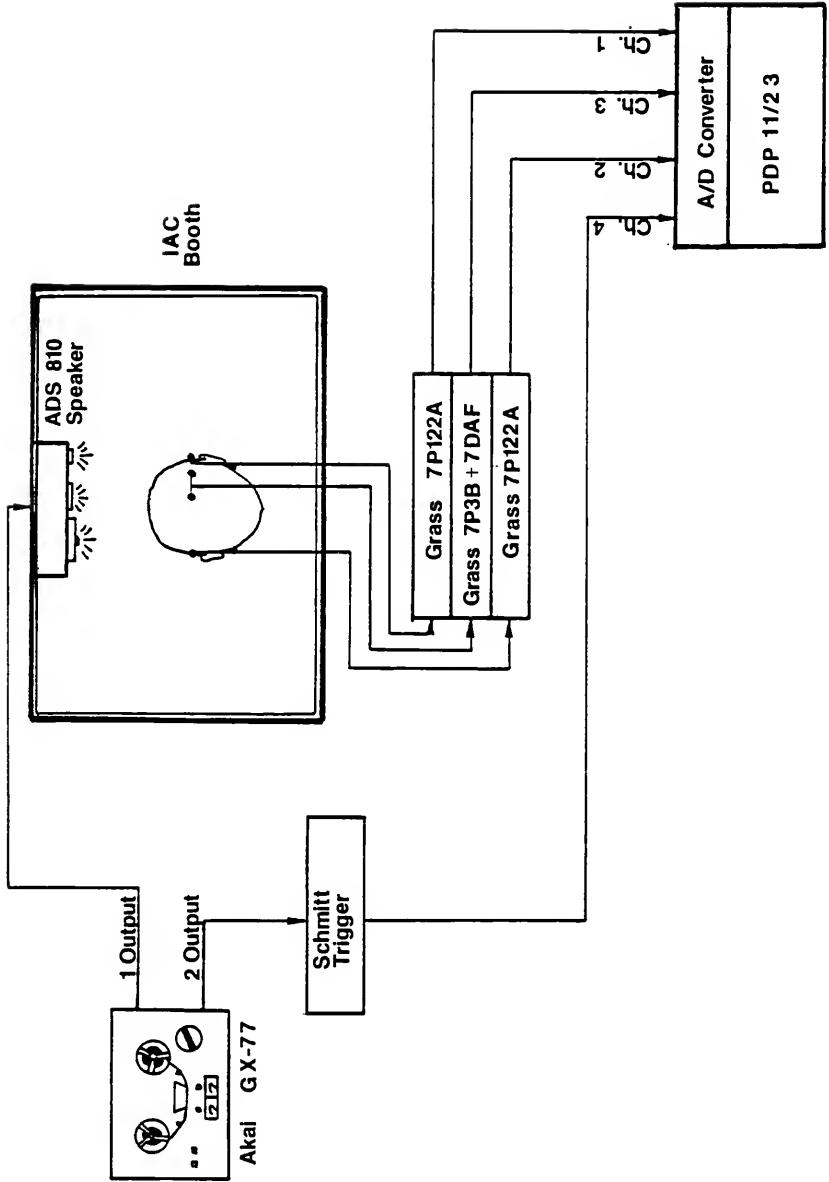
determined during a pilot study such that eyeblinks and facial movements resulted in amplified potentials which did not exceed the limits of the amplifier, but were measurably greater than ongoing facial electrical activity.

Analog to digital conversion. Immediately following amplification, the unprocessed EEG waves from the three amplifiers were input to three channels of an analog to digital (A/D) conversion device and Digital Equipment Corporation PDP 11/23 computer. The Schmidt trigger pulse was fed into a fourth channel of the A/D board. At the occurrence of each pulse, corresponding to the onset of each stimulus item, the electrocortical waves in channels one through three were digitized at the rate of 200 Hz (one sampling every 5 ms) for a period of 500 ms, resulting in 100 voltage values per wave. These digitized waves were stored on hard disk for later averaging on the PDP 11/23. Channel three, which received input from electrodes placed near the eye, was utilized as an artifact rejection channel. When the absolute voltage of channel three exceeded 1.8 V, indicating an eyeblink or facial muscle movement, the data on channels one and two were dropped, and thus were not available for later averaging. See Figure 2-3 for the equipment configuration.

#### Preliminary Data Analysis

Extraction of AER's. The subject's perceptual responses were scored at the conclusion of each experimental

Figure 2-3. Equipment configuration for stimulus presentation during the electrocortical recording procedure.



session. At this juncture, a selective averaging program was utilized; it first allowed the digitized response waves associated with incorrect responses to be excluded from the averaging process. This procedure was carried out for both the natural and synthetic syllable trials, on the assumption that it would maximize differences between AER's associated with /b/ and /d/. If more than five responses for any syllable had to be eliminated, due to incorrect perception and/or muscle artifacts, the subject was tested on that particular trial a second time. For the synthetic syllable trial and the natural speech trial, an average of 19.5 responses and 18.7 responses per syllable, respectively, were available as a basis for obtaining each AER.

Subjects incorrectly identified approximately half the chirps in the two chirp trials, so elimination of all incorrect responses was not possible. As a result, all chirp responses not contaminated with muscle artifacts were included in the averaging process. For the chirp trials, an average of 19.7 responses per syllable was utilized in obtaining each AER. The procedures described above resulted ultimately in 576 separate AER's based on 12 subjects, 4 trials, 2 consonants, 3 vowels and 2 hemispheres.

Normalization of AER's. Due to equipment limitations, precise calibration of the biological amplifiers was not possible. They were calibrated prior to each use with a 1 mV square wave pulse, but calibration in microvolts could

not be accomplished because measuring devices of adequate sensitivity were not available. Two methods were employed to compensate for this limitation and insure that the amplifiers associated with T3 and T4 were equivalent. First, the two 7P122A amplifiers were balanced over hemispheres and conditions so that for each condition, half the responses from a particular hemisphere were amplified by one amplifier and half by the other. A second manner in which potential amplifier differences were eliminated from the data was by normalizing each AER. This process was carried out by converting each of the 100 voltage points to Z-scores--a procedure which involves subtracting the mean of all points of a particular wave from each individual point and dividing by the standard deviation (i.e.,  $Z=(x-\text{MEAN})/\text{STANDARD DEVIATION}$ ). This procedure had the effect of aligning all the AER's along a common baseline and equalizing peak amplitudes. Once normalized, the entire data set was sent via modem to an Amdahl 470 V/6-11 computer in the University of Florida's Northeast Regional Data Center for statistical processing.

Analysis of the average evoked responses. Analysis of waveforms comprising AER's traditionally has been a difficult task, due to the complexity of the response. Thus, the question arises: what procedures can be used to measure a waveform of this type? Several methods have been



frequently utilized by AER researchers, including various types of peak measurement and area analysis.

Peak analysis is based on the assumption that it is only necessary to measure the waveform at a limited number of points in relating electrophysiological response with cognitive variables. Although peak analysis of individual responses is intuitively appealing and does not require sophisticated computer interface, it has a number of disadvantages. First, peak identification is dependent on experimenter interpretation. Due to variations in latency, and the frequent presence of several peaks at the desired latency, the experimenter must often make subjective decisions relative to the precise point at which measurement should be made. Second, there may be a large number of individual AER's to be analyzed, depending on the number of subjects and independent variables. Since peak measurements are made by hand, the time and effort required for this type of analysis may be prohibitive for the more complex experimental designs. A third serious disadvantage with this technique is the necessary assumption that the peaks observed in the waveform are independent, and not caused by some single underlying process. Finally, additional technical problems, such as reliable estimates of baseline and spurious values at the point being measured, reduce the utility of this approach.

Some experimenters, such as Wood (1975), have utilized peak analysis on grand mean AER's rather than on individual waveforms. A "grand mean AER" is a composite waveform derived from averaging responses over all subjects for a particular experimental condition. This technique has the advantage of producing smooth waveforms with easily defined peaks, since a large number of individual AER are generally averaged in calculating each grand mean AER's. Further, the averaging can be done by computer, and results in several composite AER's rather than hundreds of individual waveforms, thus simplifying the final peak and latency measurements which are done by hand. However, the problem of reliable baseline estimates remains. In addition, comparisons between group averaged AER's do not take into account inter-subject variability; thus, comparing peaks (or all points comprising a wave, as Wood, 1975 did) for significant differences may produce inaccurate results due to large variances in the data.

Area measurements overcome some of the disadvantages of peak analysis, but this technique also is somewhat limited. In this case, amplitude measures within a latency range of interest are integrated; hence the measure is less subjective than peak measurement and less subject to spurious values. However, a number of disadvantages exist. It is not possible to specify the underlying components present in the wave, and how these components may relate to

multidimensional experimental variables. Further, integration limits must often be set arbitrarily because the experimenter does not know the location of the underlying components, and baseline estimates continue to be a problem.

The concept of "underlying components" is an important one when considering AER measurement techniques. According to Donchin, Ritter and McCallum (1978), most researchers consider the individual peaks comprising their observed waveforms as "components." However, as Donchin et al. (1978) point out, it is more probable that the observed AER waveform is the sum of a number of underlying "component waves," which occur both sequentially (in serial) and simultaneously (in parallel). These authors define components as reflecting "the activity of . . . functionally distinct neuronal aggregates" (pg. 5). Thus, "components" are hypothesized to represent specific neural processes which occur in response to particular aspects of a stimulus.

In any case, it is possible that these cited component waves vary reliably as a function of experimental manipulations, and result in a more complete description of cognitive processing than peak or area analysis reveals. Chapman, McCrary, Bragdon and Chapman (1979) furnished support for this theory, by relating underlying components extracted through Principal Component Analysis to various aspects of information-processing tasks. Their results

revealed two components which correlated with previously identified surface phenomena, the contingent negative variation (CNV) and a late positive peak (P300). These two features were associated with expectancy of relevant stimuli and the presentation of relevant stimuli, respectively. However, Chapman et al. were also able to isolate additional AER components correlating with other processing tasks which had not been previously noted. Thus, it appeared that Principal Component Analysis allowed a more complex analysis of the effects of experimental variables than traditional measurement techniques. For this reason, Principal Component Analysis (PCA) was chosen for use in the present research.

According to Donchin and Hefley (1978), there are several disadvantages to be considered in applying PCA to AER research. First, it is not intuitively obvious how the PCA values relate to the original waveforms, and the experimental results may be difficult to interpret. A more serious flaw in terms of data analysis is that PCA is not resistant to artifacts created by variations in peak latency. Amplitude differences at a particular latency are treated as if all waves peaked at the same point in time, which may or may not be the case. Potentially, this disadvantage is overcome by use of careful recording techniques, by examination of the data prior to PCA application and by adjustment of latencies if necessary.

Chapman et al. (1979) did not appear to consider this latency variation a problem in extracting components and reconstructing original AER's, and other researchers using this technique have not mentioned latency variation as a problem prior to analysis or a confounding factor post hoc (Donchin et al., 1978; Molfese, 1978a; Molfese, 1978b; Molfese, 1980a; Molfese and Schmidt, 1983).

An additional consideration when applying this type of analysis is the lack of physiological evidence to support the validity of components. Although "neuronal aggregates" have been hypothesized as the source of these factors, such structures have not been isolated in the cortex. PCA is a mathematically parsimonious procedure, which isolates components solely on the basis of correlations, axis rotations, and other formulae. Proponents of PCA, such as Donchin et al. (1978), would argue that the strong relationship between a component and an experimental variable can furnish important information about cognitive processing, regardless of the source of the component. That point of view is adhered to in this study.

#### The Principal Component Analysis Procedure

The AER waveform can be conceptualized as a series of voltage measurements; and the "variables" in PCA are these voltage values. The number of variables in any given study is determined by digitization rate of the computer and

duration of the averaging epoch. For example, in the present study, the sampling rate was 200 Hz for 500 ms, resulting in 100 voltage values, or variables, for each of the 576 AER's. Thus, each waveform was represented as a series of 100 discrete numbers.

#### Calculating the Centroid

The first step in PCA is to average each variable over all AER's in the data set. This procedure results in a grand mean AER known as the centroid. In turn, the centroid reflects the average voltage value at each time point for all AER's. This measure is used as the basis for factor extraction.

#### Matrix Construction

The next step in PCA is to construct a matrix in which all the voltage values are correlated with each other. If the raw data are used, this matrix is referred to as a "cross-products" matrix. In this case, the total variance of the data set is analyzed. Alternatively, a "covariance matrix" can be used, in which the mean of all the voltage values at a particular time point are subtracted from each original AER at that time point, prior to computing the matrix. This procedure has the effect of removing that portion of the variance due to differences in means. Finally, a "correlation matrix" may be used, in which the mean of all values at a certain time point is subtracted from each original AER (as in the covariance matrix) and the

remainder is divided by the standard deviation of all voltage values at that particular time point. The result of this treatment is to normalize peak amplitudes over all AER's.

According to Donchin and Heffley (1978), use of the covariance matrix is most desirable in AER research. The cross-products matrix may result in components related more to subject variability than to experimental manipulation, and the correlation matrix may give too much weight (due to normalization) to small, unreliable differences in waveforms. Analysis of the covariance matrix is based on the difference between an individual AER and the grand mean, and this is most useful when an analysis of the effects of experimental manipulations across subjects is planned.

#### Extraction of Principal Components

Following the calculation of the centroid and the construction of the matrix, the next step in PCA is to extract the principal components or factors. (In this study, the terms "components" and "factors" will be used interchangeably, as they are in current AER literature; however, according to Donchin and Heffley, 1978, the label "component" is correct). Factor extraction involves reduction of the variables in the matrix to a predetermined number of linear combinations, or factor loadings, which account for the most possible variance in the data. Each factor loading consists of  $n$  coefficients corresponding to

the original time points, and reflects the influence of each factor (component) on that time point.

#### Uncorrelating the Factors

The next step in PCA is to rotate these factor loadings in order to maximize orthogonality. When attempting to relate underlying components (or factors) to experimental variables, it is desirable to have each factor as uncorrelated as possible with other factors. Since the initial factors extracted from the centroid tend to be somewhat correlated (due to the sequential nature of the variables), some type of rotation is necessary in order to improve orthogonality. Varimax rotation (Kaiser, 1958) is traditionally used. The result of this procedure is to concentrate the high loadings for each factor within a given time range, thus producing distinct AER components.

#### Derivation of the Factor Scores

The final step in PCA is to transform the original AER's to the new, rotated axes. This transformation is accomplished by multiplying each original AER by a coefficient vector derived from the rotated factor loadings. A number of factor scores, equal to the number of factors, is the result of this process. These factor scores represent a measure of the magnitude of a specific factor in a particular AER. Factor scores (for the factor being analyzed) can then be averaged over experimental conditions to yield a mean factor score, which in turn can be utilized



as the dependent variable in an Analysis of Variance. In this way, the effect of experimental manipulations on electrocortical activity can be assessed.

In this study, a separate Analysis of Variance (ANOVA) was calculated for each factor. Mean factor scores were compared between levels of the independent variables Consonant, Vowel, Hemisphere and Trial in each ANOVA. Following this assessment of main effects, variables were compared in all possible combinations for two-, three-, and four-way interactions.

## CHAPTER III RESULTS

### Preliminary AER Data Analysis

The electrocortical recording procedure utilized in this research resulted in a total of 576 separate AER's. This value was obtained from 12 subjects responding to two consonants in combination with three vowels from both hemispheres in four separate trials ( $12 \times 2 \times 3 \times 2 \times 4 = 576$ ). Two examples of the unprocessed AER's are presented in Figure 3-1. Each waveform is based on approximately 20 repetitions of the syllable /bi/, and each is from a different subject. The AER's then were normalized, as described above, and subjected to off-line Principal Component Analysis (PCA). Finally, the output of this preliminary statistical procedure was utilized in ten Analyses of Variance (ANOVA'S), and subsequent preplanned and post hoc comparisons. The PCA and ANOVA procedures were carried out twice: once on the full data set of AER's, and a second time on only the AER's associated with the synthetic and natural syllables. Finally, the perceptual results of this study were analyzed.

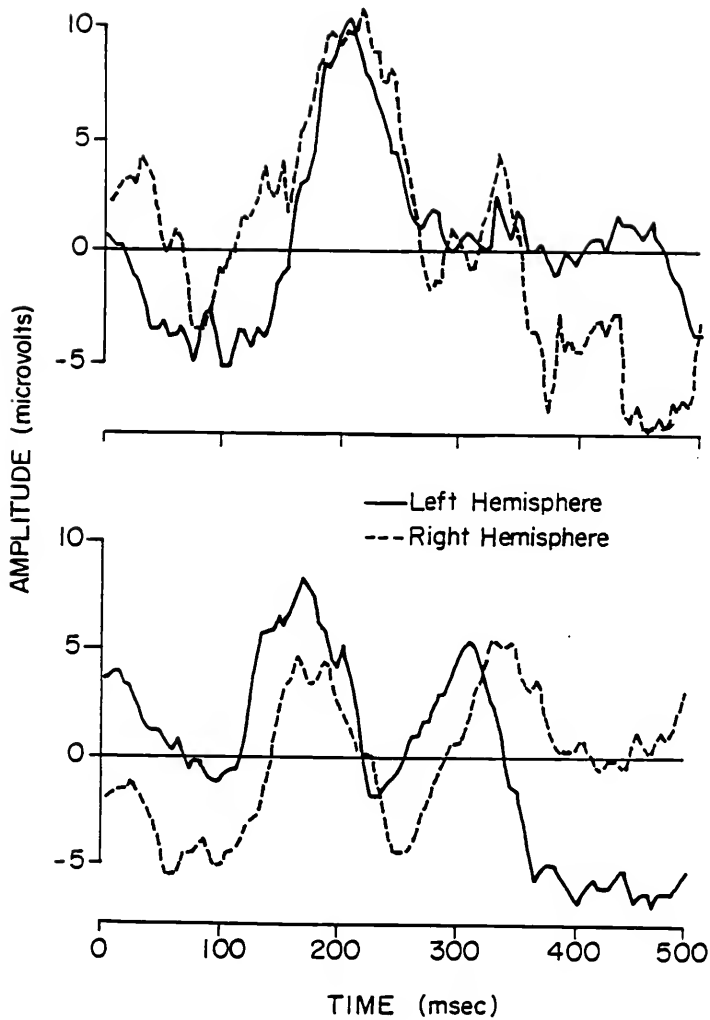


Figure 3-1. Normalized AER's based on approximately 20 repetitions of the syllable /bi/ from (a) subject 1 and (b) subject 2.

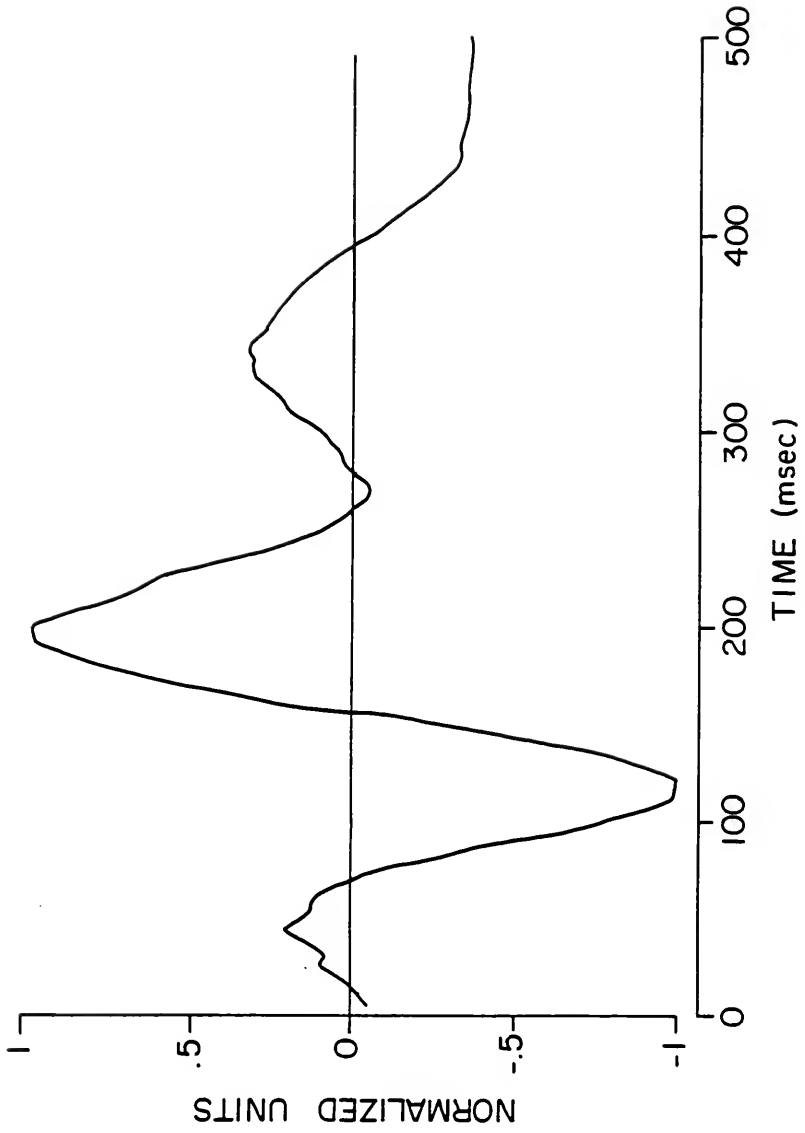
Analysis One: The Full Data Set

The reasons for analyzing the full data set were (in part, at least) to evaluate the findings of Molfese (1980a) and Molfese and Schmidt (1983) regarding left hemisphere differentiation of stop consonants. A further goal was to provide electrophysiological evidence for the perceptual changes noted with differences in stimulus expectation (Schwab, 1981; Nusbaum et al., 1983). Statistical procedures utilized in investigating these issues were PCA and ANOVA's, as well as preplanned and post hoc comparisons.

The first step in the PCA was to calculate the centroid, or average, of all 576 normalized AER's (Dixon, 1981). The centroid is pictured in Figure 3-2. It is characterized by a small positive peak at 45 ms (P45), a large negative peak at 120 ms (N120), a large positive peak at 195 ms (P195), a negative peak at 270 ms (N270), a small positive peak at 340 ms (P340), followed by a gradual--and negative--decline asymptoting at 455 ms (N455). This centroid is very similar in waveshape to the one reported by Molfese and Schmidt (1983), who showed a P30, N120, P200, N270, P345, and N450. The main difference occurred in the final 150 ms of the wave, during which the present study found a falling configuration while Molfese and Schmidt (1983) found a level to rising configuration.

The next step in the PCA was formation of a 100 x 100 covariance matrix and extraction of the principal components

Figure 3-2. The centroid, or grand mean auditory evoked response obtained from the Principal Components Analysis based on the full data set (Analysis One).



(or factors). Factors with eigen values of one or more were retained for further analysis (Chapman et al., 1979). This procedure resulted in 10 factors which accounted for 62.7% of the variance. Factors then were rotated using a varimax criterion (Kaiser, 1958) in order to improve orthogonality. After 14 iterations, the terminal solution was reached. The rotated factors are pictured in Figures 3-3. These factors, or component waves, are assumed to underlie the surface waveshape of the centroid, and to be present to a greater or lesser extent in each individual AER. Peaks in these factor waveshapes represent the latency at which a specific factor affected the centroid, regardless of polarity (Molfese and Schmidt, 1983). Factor 1 was characterized by a positive peak at 40, a negative peak at 90 ms, and a major positive peak at 150 ms. This component influenced the centroid at P45 and the N120-P195 complex. Factor 2 was characterized by a positive peak at 75 and a small negative peak at 145 ms; it influenced the centroid at the P45-N120 complex. Factor 3 showed a major peak at 25 ms, a small negative peak at 85 ms and a positive peak at 120 ms, and also influenced the P45-N120 complex of the centroid. Factor 4 had several small peaks throughout its duration and one major peak at 200 ms. This major peak influenced the P195 of the centroid. In Factor 5, a major positive peak occurred at 330 ms followed by a small positive peak at 440 ms. This factor probably influenced N340 and the declining latter

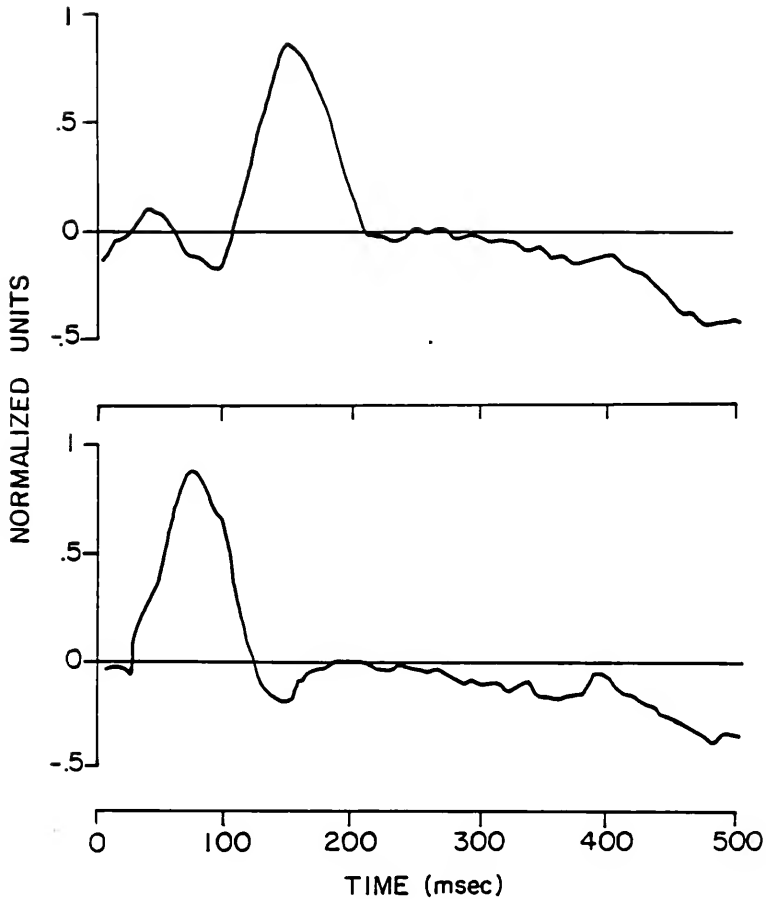
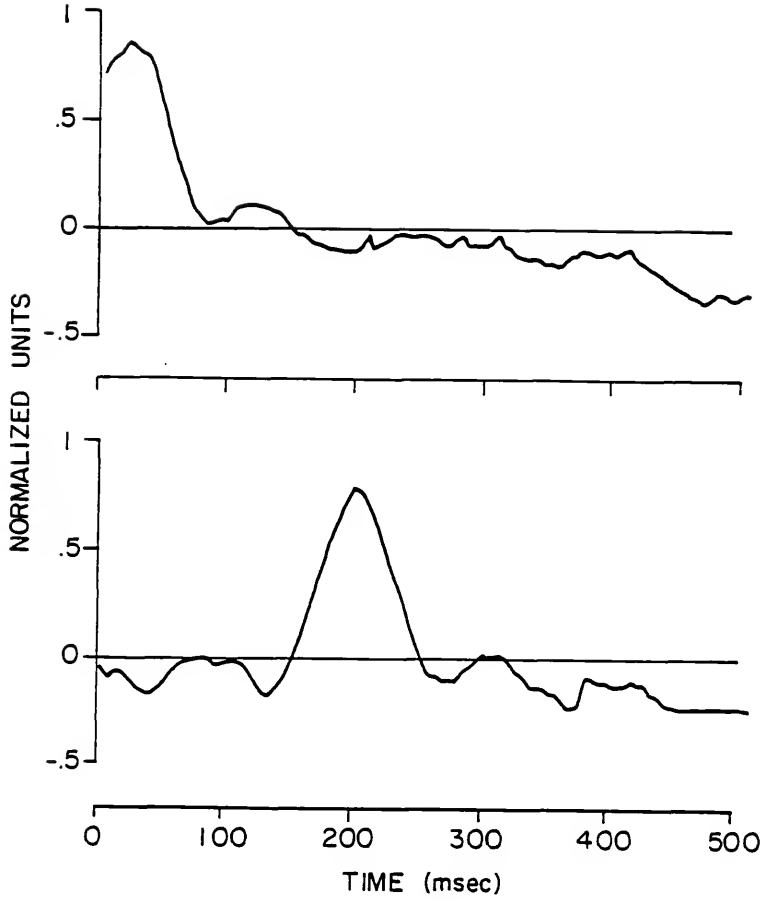
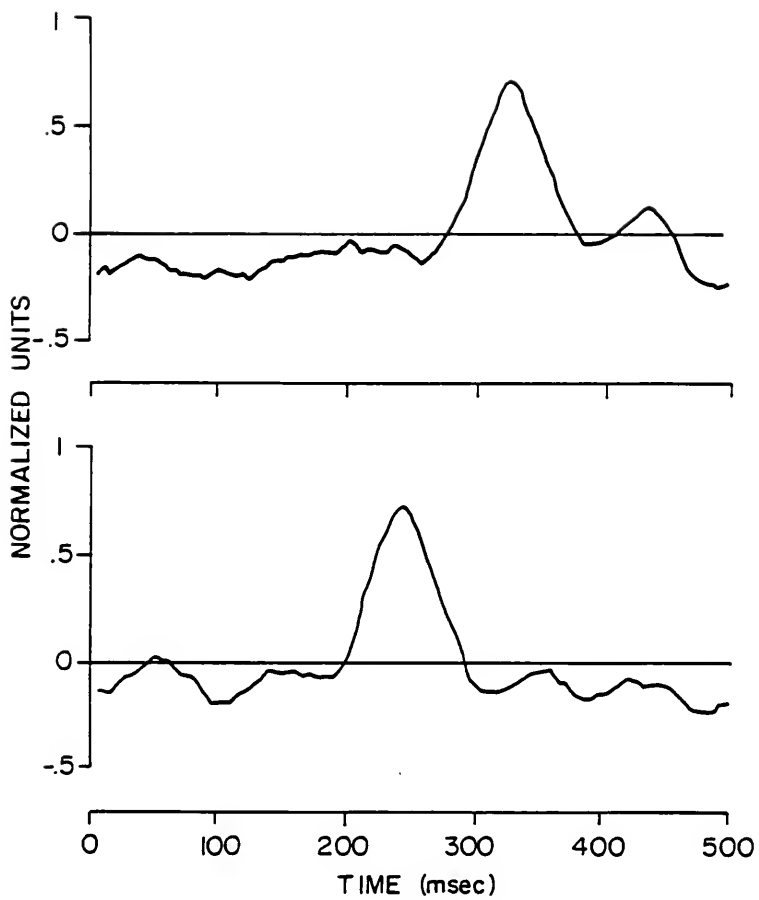
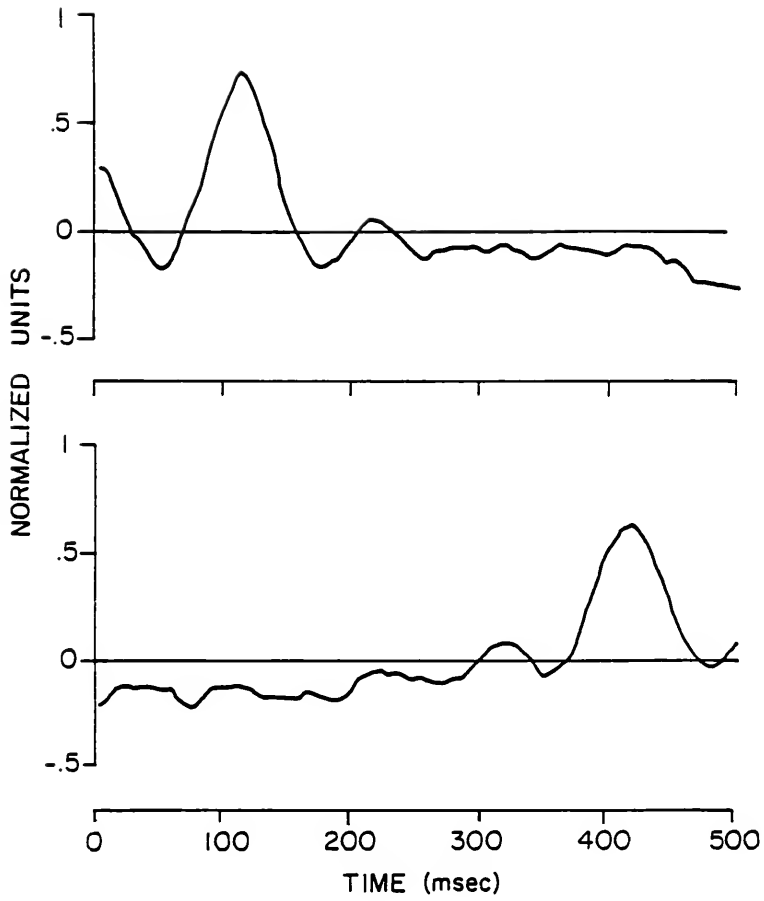


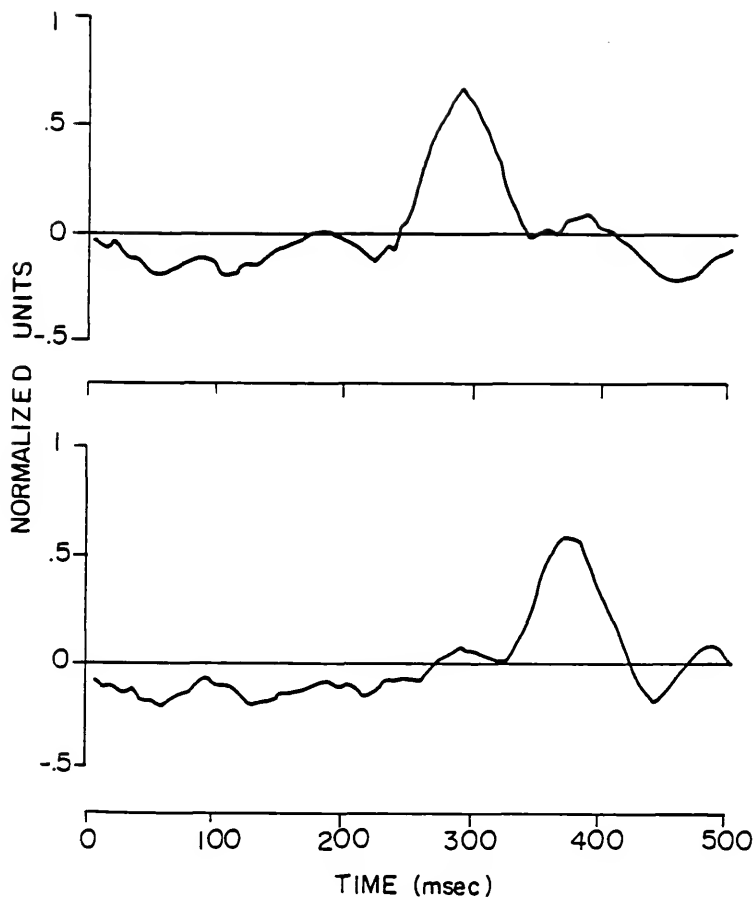
Figure 3-3. The ten factors extracted by means of a Principal Components Analysis based on the full data set (Analysis One).  
 (a) Factor 1, (b) Factor 2, (c) Factor 3,  
 (d) Factor 4, (e) Factor 5, (f) Factor 6,  
 (g) Factor 7, (h) Factor 8, (i) Factor 9,  
 (j) Factor 10











part of the centroid. Factor 6 was characterized by several small peaks, similar to Factor 4, with a major peak at 245 ms. This component probably influenced the P195-N270 portion of the centroid. Factor 7 showed a positive peak at 0 ms, a negative peak at 50 ms, a major positive peak at 115 ms, a small negative peak at 175 ms and a small positive peak at 220 ms. This factor appeared to have its major influence at the P0-N120 portion of the centroid. For Factor 8, a small positive peak at 320 ms followed by a major positive peak at 420 ms influenced the P340-N455 complex of the centroid. Factor 9 contained a major positive peak at 290 ms, a small positive peak at 385 ms and a negative peak at 450 ms and influenced the portion of the centroid just after N270. And finally, Factor 10 was characterized by a major positive peak at 375 ms, a negative peak at 445 ms and a small positive peak at 480 ms, influencing the final epoch of the centroid.

The final step in the PCA was calculation of ten sets of factor scores for each of the original 576 AER's (based on the ten extracted factors). Thus, each AER in the data set was effectively represented by ten factor scores in place of its original 100 voltage values.

At this point, factor scores for each AER were utilized as the dependent variables in ten separate ANOVA's (one for each factor). All possible main effects and interactions

for the independent (classification) variables of Consonant, Vowel, Hemisphere and Trial were calculated (Dixon, 1981).

In assessing the significance of ANOVA results, a probability of .05 was chosen, in order to include as many main effects and interactions as possible while maintaining a reasonably high level of significance. The .05 level is appropriate when the data are being explored for significant trends in new research areas. The .01 level was considered too stringent, with too great a possibility of rejecting major effects and interactions (Type II error).

#### Primary Hypothesis Analysis

The principle question addressed in Analysis One was whether /b/ and /d/ were differentiated in the left hemisphere for trials which included both syllable stimuli and ambiguous stimuli (chirps with speech instructions), essentially a replication of Molfese (1980a) and Molfese and Schmidt (1983). Such a finding would support a hypothesis of left hemisphere involvement in the perception of voiced stop consonants. In order to test this relationship, the ten ANOVA's described above were examined for significant Consonant by Hemisphere by Trial interactions.

The ANOVA of one factor (Factor 9) did indeed reveal a significant Consonant by Hemisphere by Trial interaction ( $F = 3.63$ ,  $p = .0229$ ,  $df = 3,33$ ). However, this result is somewhat ambiguous as the interaction contained 16 mean factor scores, obtained from two consonants by two

hemispheres by four trials. Although the entire interaction was found to be significant, it was not apparent which pairs or combinations of means were significantly different. Thus, in order to specify significant combinations of mean factors scores, post hoc testing was necessary.

A t-square Planned Comparison procedure was utilized in the post hoc analysis of the significant Consonant by Hemisphere by Trial interaction. Mean factor scores of /b/ in the left hemisphere were averaged over the three trials, and compared with those associated with /d/. For this test, a probability level of .01 was chosen, in order to reduce the possibility of concluding that differences were significant when in fact they were not (Type I error). This more conservative level was considered necessary because a t-square Planned Comparison does not control error rate simultaneously for multiple comparisons, and thus repeated tests on the same set of data greatly increase the chances of a Type I error. Such a statistic is appropriate only for planned comparisons when ten or fewer comparisons are being made, at significance levels of .01 or better, according to Shearer (1982). Results of this comparison revealed that when /b/ and /d/ were compared in the left hemisphere, averaged over synthetic syllables, natural syllables and chirps with speech instructions (speech chirps), differences between means failed to attain significance at the .01 level (although  $F = 7.02$ ,  $p = .0118$ ). Thus, a hypothesis of left

hemisphere involvement in the perception of voiced stop consonants was not supported. This result was not consistent with the findings of the Molfese research for the consonants /b/ and /g/.

A similar post hoc comparison between right hemisphere means was also made for this Consonant by Trial by Hemisphere interaction. This procedure was carried out in order to eliminate the possibility of right hemisphere differentiation of /b/ and /d/, a phenomenon not previously reported. In this case, differences between mean factor scores were not significant, as expected ( $F = .204$ ,  $p = .659$ ).

Finally, post hoc t-square comparisons of the means in this interaction were utilized to determine whether patterns of hemispheric involvement could be changed as a function of instructions to the subjects (stimulus expectation). In this procedure, the mean factor score for "low" stimuli (identical to /b/) in the frequency chirp trial was compared to the mean factor score for the "high" stimuli (identical to /d/). Comparisons were made for both the left and the right hemispheres. Results revealed that /b/ vs. /d/ in the left hemisphere were, again, not significant as expected ( $F = .681$ ,  $p = .580$ ). When a similar test was made in the right hemisphere, differences between /b/ and /d/ did not attain significance at the .01 level (although  $F = 6.77$ ,  $p = .0131$ ). Based on these data, it appeared that changes in

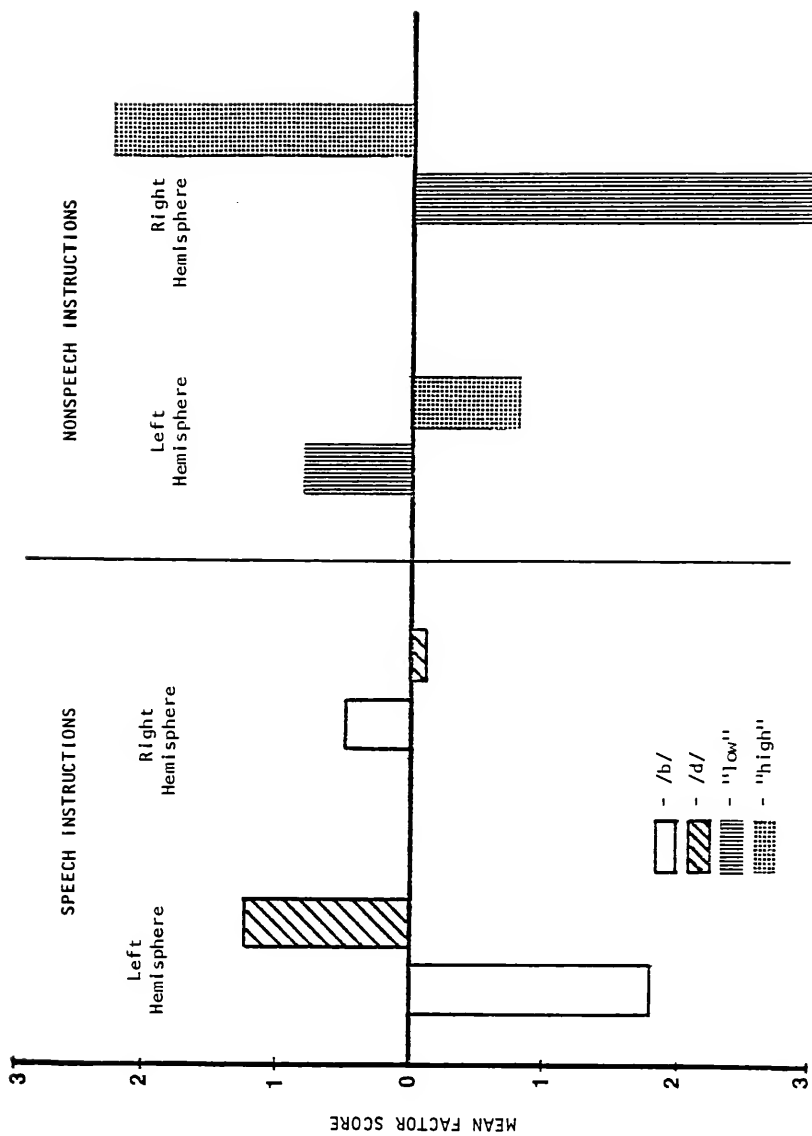


instructions did result in some shift in hemispheric involvement, although this trend was not significant.

A graphic display of mean factor scores in the four comparisons for this interaction is presented in Figure 3-4. The first set of bars illustrates the difference in mean factor scores for /b/ vs. /d/ in the left hemisphere averaged over the natural syllable, synthetic syllable and speech chirp trials. The difference between /b/ and /d/ is substantial, although not significant at the .01 level. The second set of bars shows mean factor score differences for the same set of variables in the right hemisphere. As can be seen, the differences are negligible. Taken together, these two sets of bars display a definite trend toward left hemisphere involvement in /b/-/d/ discrimination, although this trend was not statistically significant.

The third set of bars in Figure 3-4 illustrates mean factor score difference between "low" vs. "high" judgments in response to chirps with frequency instructions (frequency chirps) in the left hemisphere. A small difference appears to occur. The last set of bars shows mean factor score for the same discrimination task in the right hemisphere. Here, a substantial difference between "low" and "high" responses can be seen, although again, this difference did not attain significance at the .01 level. The third and fourth sets of bars in this figure display a trend toward right hemisphere

Figure 3-4. Mean factor scores for /b/ vs. /d/ in the left and right hemispheres; and "high" vs. "low" onsets in the left and right hemispheres. See text for discussion.



involvement in frequency discrimination--although again, this trend was not significant.

A visual inspection of the "Speech Instructions" portion of Figure 3-4 and the "Frequency Instructions" portion reveals a tendency toward differential hemispheric involvement depending on the instructions to the subjects. Although neither relationship was significant, speech instructions appeared to result in greater left hemisphere differentiation, while frequency instructions yielded greater right hemisphere differentiation.

#### Secondary Analyses

Further post hoc analyses were undertaken on the ANOVA data in order to answer a number of additional questions. Because the data analysis at this point was exploratory in nature, Scheffe post hoc comparisons were used in determining significance of results. Significance level was set at .05 in order to reduce the possibility of making a Type II error while keeping the probability of a Type I error at a reasonable level.

The questions investigated in the secondary analyses are as follows:

- 1) Are there significant differences between responses to /b/ vs. /d/, independent of other variables? Both Molfese (1980a) and Molfese and Schmidt (1983) found such a relationship, which they interpreted as a bilateral process

differentiating between stop consonants. In the present study, such a process might also be expected.

Results revealed that a main effect for Consonant characterized Factor 1, significant at the  $p = .0021$  level ( $F = 15.98$ ,  $df = 1,11$ ). Because this was the only comparison possible for this term, and because it was significant, no other analysis was undertaken. Thus, Molfese's hypothesized bilateral process was supported by the present research. Further, the latency of the major peak associated with Factor 1 (150 ms) was in general agreement with the latency reported by Molfese for this bilateral differentiation (170 ms).

2) Do the left and right hemispheres appear to function differently, independent of other variables? The results of the Molfese research indicate that this is indeed the case. Such an effect was also hypothesized in the present study. Failure to find significant differences between the two hemispheres independent of other variables would confuse interpretation of hemispheric asymmetry found in higher-order interactions. The results, however, revealed a main effect for Hemisphere for Factor 4, significant at the  $p = .0293$  level ( $F = 6.27$ ,  $df = 1,11$ ). Because this term contained only two means, and comparison between them was significant, no further analysis was carried out. This finding of significant differences between the left and right hemispheres is consistent with previous research and

the concept of differential hemispheric functioning during speech perception.

3) Are there bilateral processes which differentiate between vowels, independent of other variables? Molfese and Schmidt (1983) found a number of bilateral processes which appeared to permit discrimination between various pairs and groupings of their three vowels. Similar bilateral processes were hypothesized in the present study. In order to begin testing this assumption, the ANOVA results were explored for significant vowel main effects.

Four factors (2, 4, 7 and 9) were characterized by significant vowel main effects. ( $F = 9.40$ ,  $p = .0011$ ;  $F = 4.60$ ,  $p = .0214$ ;  $F = 3.75$ ,  $p = .0398$ ;  $F = 7.31$ ,  $p = .0037$ ;  $df$  for all = 2,22). However, since each main effect contained three mean factor scores (one for each vowel), post hoc testing was necessary in order to determine which pairs or groups of vowels were significantly different.

Post hoc analyses of Factor 2 (major peak latency at 75 ms) revealed significant differences between /i/ and /æ/, /i/ and /ɔ/, and between /i/ vs. /æ, ɔ/ at the .05 level. Factor 7 (major peak latency at 115 ms) also revealed /i/ and /ɔ/ differentiation bilaterally, as well as /i/ and /æ, ɔ/. Factor 4 (major peak latency at 200 ms) revealed quite a different pattern. Analyses of the vowel main effect revealed a significant difference between /æ/ and /ɔ/. No other pairwise or grouped means were significant.

Finally, for Factor 9, Scheffe analyses revealed significant differences between /i/ and /ɔ/, /æ/ and /ɔ/, and between /i, æ/ vs. /ɔ/.

These results indicated that vowels indeed were differentiated by a number of bilateral processes. At the earlier post-stimulus onset latencies (75 and 115 ms), it appeared that /i/ was differentiated from the other two vowels, /æ/ and /ɔ/. Somewhat later (200 and 290 ms), it appeared that /æ/ and /ɔ/ were discriminated from each other. (The /i, ɔ/ combination was also differentiated at this latency). This difference in latencies may be due to ease of discrimination: the spread vowel /i/ may have been very easily separated from the more close vowels /æ/ and /ɔ/; while the more difficult discrimination between /æ/ and /ɔ/ took place somewhat later. More research appears needed to test these relationships.

4) Can natural syllables, synthetic syllables, chirps with speech instructions and chirps with frequency instructions be differentiated on the basis of cortical response? Previous research utilizing synthetic CV's vs. sinewave formant CV's (Molfese and Schmidt, 1983) revealed bilateral (and unilateral) discrimination. It was hypothesized that a similar bilateral process would be found in the present study, although different types of stimuli were used.

In order to test this assumption, the ANOVA data were scanned for Trial main effects. Such relationships were found for Factors 1, 2, 4, 8 and 10 ( $F = 19.62, p = .0000$ ;  $F = 30.05, p = .0000$ ;  $F = 14.12, p = .0000$ ;  $F = 3.87, p = .0178$ ;  $F = 7.83, p = .0004$ ;  $df$  for all = 3,33). Each significant main effect contained four mean factor scores, corresponding to the averages of the four trials (natural syllables, synthetic syllables, speech chirps and frequency chirps). Therefore, in order to assess significant differences between pairs of trials, post hoc testing was necessary.

Post hoc Scheffe analyses for Factor 1 (major peak latency at 150 ms) revealed that at the pairwise level, the synthetic syllable (SS) trial was significantly different from the natural syllables (NS) trial, speech chirps (CS) trial, and the frequency chirps (CF) trial ( $p < .05$ ). The NS trial was also significantly different from the CS trial and the CF trial. On the other hand, the CS and CF trials were not significantly different. For Factor 2 (major peak latency at 75 ms), a similar pattern of significance was obtained at the pairwise level. Again, all trial comparisons except the one between CS and CF were significantly different.

Results were somewhat different when Factors 4, 10 and 8 were considered. For both Factors 4 and 10, post hoc Scheffe analyses at the pairwise level revealed



nonsignificant differences between the NS and SS trials. The differences between the CS and CF trials were also insignificant. Thus, the NS and SS trials were significantly different from the CF and CS trials at the latencies associated with Factors 4 and 10 (200 ms and 375 ms, respectively). Finally, results of post hoc Scheffe analyses at the pairwise level for Factor 8 (major peak latency at 420 ms) revealed that only the SS and NS trials were significantly different at the .05 level.

Based on the data analysis for Trial main effects, a sequence of information-processing steps for different stimulus types can be hypothesized. Initially, stimuli appeared to be discriminated based on gross acoustic characteristics. That is, all the stimulus types--natural syllables, synthetic syllables, and chirps--were differentiated during the 40-150 ms interval post-stimulus onset. The speech chirps and frequency chirps were acoustically identical, thus they were not discriminated at this early stage. At the second processing stage, stimuli appeared to be differentiated as a function of duration. This stage occurred 200 to 375 ms post-stimulus onset, and during this latency range, stimuli with similar durations (SS vs. NS; or CS vs. CF) were not discriminated. Finally, at 420 ms post-stimulus onset, only the natural syllables and the synthetic syllables were differentiated. This late process may also be a function of duration, and may reflect

cognitive processing of later-occurring acoustic differences between the two types of syllables, perhaps during the steady-state vowel.

Finally, it should be noted that no cortical differentiation between the CF and CS trials was obtained. This is somewhat surprising, since task requirements and presumably cognitive processing demands were not the same for the two chirp trials. One reason for this failure to find a significant difference may be that the two tasks in fact did not require dissimilar cognitive processes. An alternative explanation is that differences in cognitive processing did occur, but later than 500 ms, and thus were not included in the data. It also is possible that individual variability in cortical response was too great to allow meaningful comparisons between chirp trials. Finally, it may be that responses to different stimulus parameters (exogenous components) were more reliably extracted from the AER data than responses reflecting different cognitive processes (endogenous components).

5) Is /b/ differentiated from /d/ in the left hemisphere, independent of trial or vowel? Molfese (1980a) and Molfese and Schmidt (1983) demonstrated this effect; however, they did not include a nonspeech trial, as this research did. Thus, it was possible that such a relationship would not be found, due to the confounding effects of trial (or stimulus expectation).

The results of the ANOVA procedures revealed a significant Consonant by Hemisphere interaction for Factor 8 ( $p = .0485$ ,  $F = 4.92$ ,  $df = 1,11$ ). However, post hoc Scheffe analyses of this interaction revealed that /b/ and /d/ were not significantly different in the left hemisphere, nor were they significantly different in the right hemisphere. This result was consistent with a hypothesis of the importance of stimulus expectation in hemispheric processing.

6) Can differential hemispheric involvement be demonstrated as a function of trial? As previously mentioned, Molfese and Schmidt (1983) found an early bilateral component which differentiated between stimulus classes (synthetic syllables and sinewave formant CV's), and they also found a similar differentiation unilaterally during a later post-stimulus epoch. In the present study, unilateral discrimination between trials (i.e., stimulus classes) was also sought.

In assessing the effects of hemisphere as a function of trial, factors with significant Trial by Hemisphere interactions were selected for further analysis. Three factors--4, 6, and 10--showed such an interaction significant at the .05 or better level. To analyze these interactions, pairs of trials were compared separately for each hemisphere. For Factor 4 (major peak latency at 200 ms), the Trial by Hemisphere interaction was significant at the  $p = .0306$  level ( $F = 3.35$ ,  $df = 3,33$ ). Post hoc Scheffe

analyses revealed that for the left hemisphere, the SS trial was significantly different from the CS trial at the .05 level, and the SS trial was also significantly different from the CF trial. No significant differences in the right hemisphere were found for this interaction. Although Factors 6 and 10 also contained significant Trial by Hemisphere interactions ( $F = 3.51, p = .0258$ ;  $F = 2.98, p = .0454$  respectively;  $df = 3,33$  for both), post hoc Scheffe analyses failed to reveal significant hemispheric effects.

These results are somewhat difficult to interpret. Syllable vs. chirp stimuli appeared to be discriminated in the left hemisphere, although inconsistently. Further, the latency of this unilateral process (200 ms) was considerably earlier than two of the obtained bilateral processing stages (375 and 420 ms). From these data, it would appear that bilateral differentiation between stimulus classes as a function of acoustic parameters was the principle perceptual process involving trials. Unilateral differentiation was present, but was only inconsistently related to stimulus or cognitive variables. Further study is needed to determine the relationships indicated by these data.

6) Is there any evidence for unilateral processing of vowels in either hemisphere? Such processing has not been noted in previous research, and was not expected in this study. However, in order to determine the possibility of

hemispheric asymmetry in vowel perception, the 10 ANOVA's for each of the 10 extracted factors were scanned for a significant Vowel by Hemisphere interaction.

The ANOVA results revealed no such interaction in any of the data. To control for the possibility of a highly significant trial effect masking a simple Vowel by Hemisphere interaction, the data were also examined for Vowel by Hemisphere by Trail interactions. Factor 7 was characterized by such an interaction ( $F = 2.24$ ,  $p = .0498$ ). However, post hoc Scheffe analyses revealed no significant differences when pairs or groups of vowels were compared in either hemisphere for any particular trial. Thus, results from this study supported previous research findings of no hemispheric asymmetry in the perception of (undistorted) vowels.

#### The Relationship Between Factors and Grand Mean AER's

As previously mentioned, it is not intuitively apparent how factors and factor scores relate to observed differences in AER waves. Theoretically, the latencies of the major peaks in the factor waveform should indicate the latencies in the AER waveforms--that is, where differences occur. Thus, for example, if the major latency of the significant factor is at 100 ms, a difference between the grand mean AER waveforms for the variable in question should be apparent in the area of 100 ms. When making such comparisons, however, it should be noted that the PCA method of factor extraction

takes into account the variability of each individual AER. Thus, it is possible that some areas of difference between the grand mean AER's are not identified by PCA as being significantly different due to variance in individual data.

For the Consonant main effect described above, /b/ and /d/ were differentiated on the basis of factor scores derived from Factor (or component wave) 1. The waveshape of Factor 1 shows peaks at 40, 90 and 150 ms; thus, a visual inspection of the averaged AER's for /b/ vs. /d/ should reveal differences between waves from approximately 40 to 150 ms latency. The averaged AER's for /b/ vs. /d/ are presented in Figure 3-5. Each wave was based on 288 separate AER's averaged over vowels, hemispheres and trials. Visual inspection revealed that the waveforms were somewhat different over the entire time course, with the most marked differences occurring at 0 to 15 ms, 75 to 100 ms, 155 to 190 ms, and 290 to 300 ms. These latencies were in general agreement with the peak latencies characteristic of Factor 1, with the exception of the 290 to 300 ms difference noted in the grand mean AER waveforms. Thus, the factor specified as differentiating between /b/ and /d/ appeared to relate to actual differences observed between grand mean AER's.

Grand mean AER's for hemispheres were plotted in Figure 3-6. Factor 4 (major peak at 200 ms) was found to be associated with significant differences between hemispheres. Visual inspection of Figure 3-6 revealed marked differences

Figure 3-5. Grand mean average evoked responses to syllables beginning with /b/ and syllables beginning with /d/.

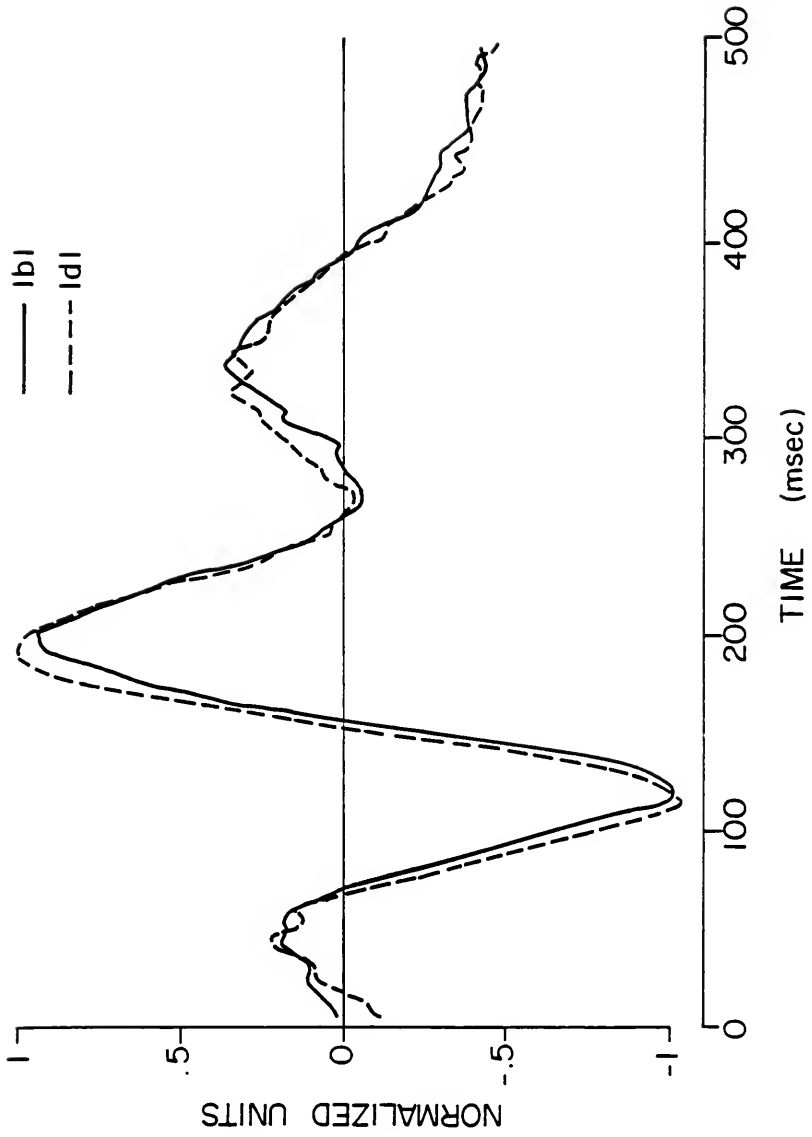
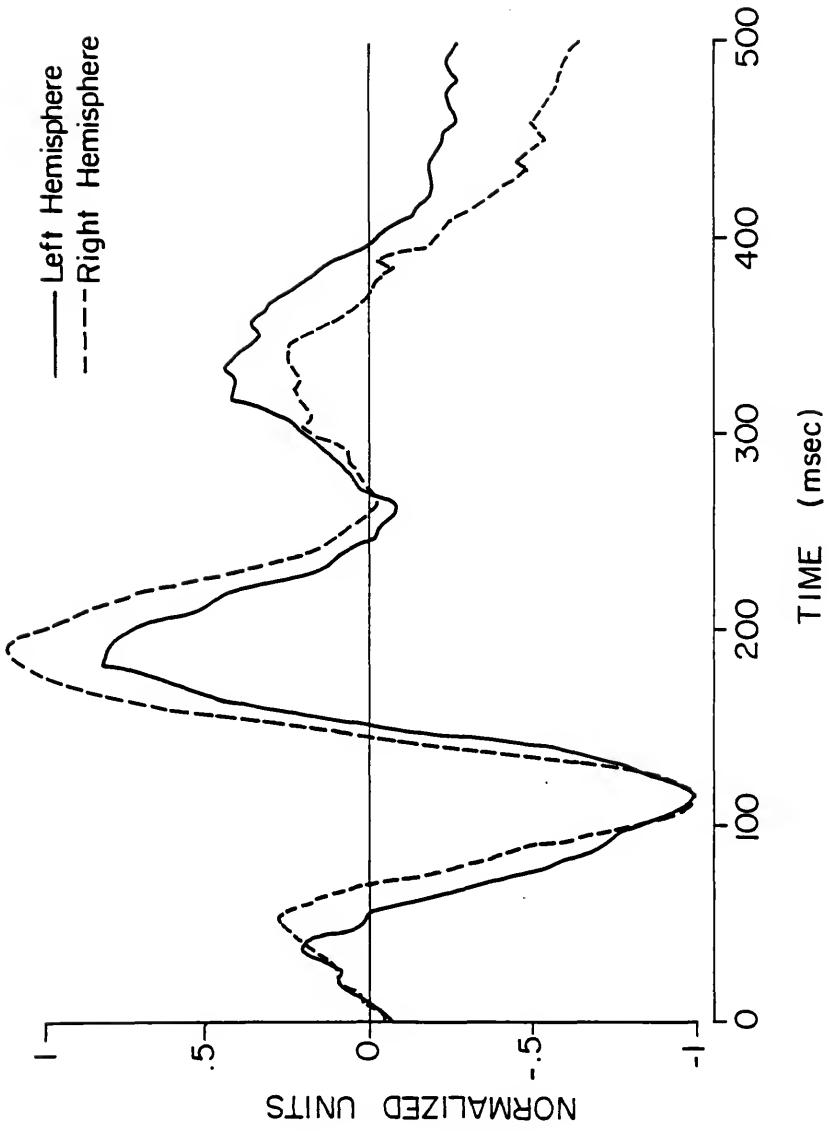




Figure 3-6. Grand mean average evoked responses from the left hemisphere and the right hemisphere.

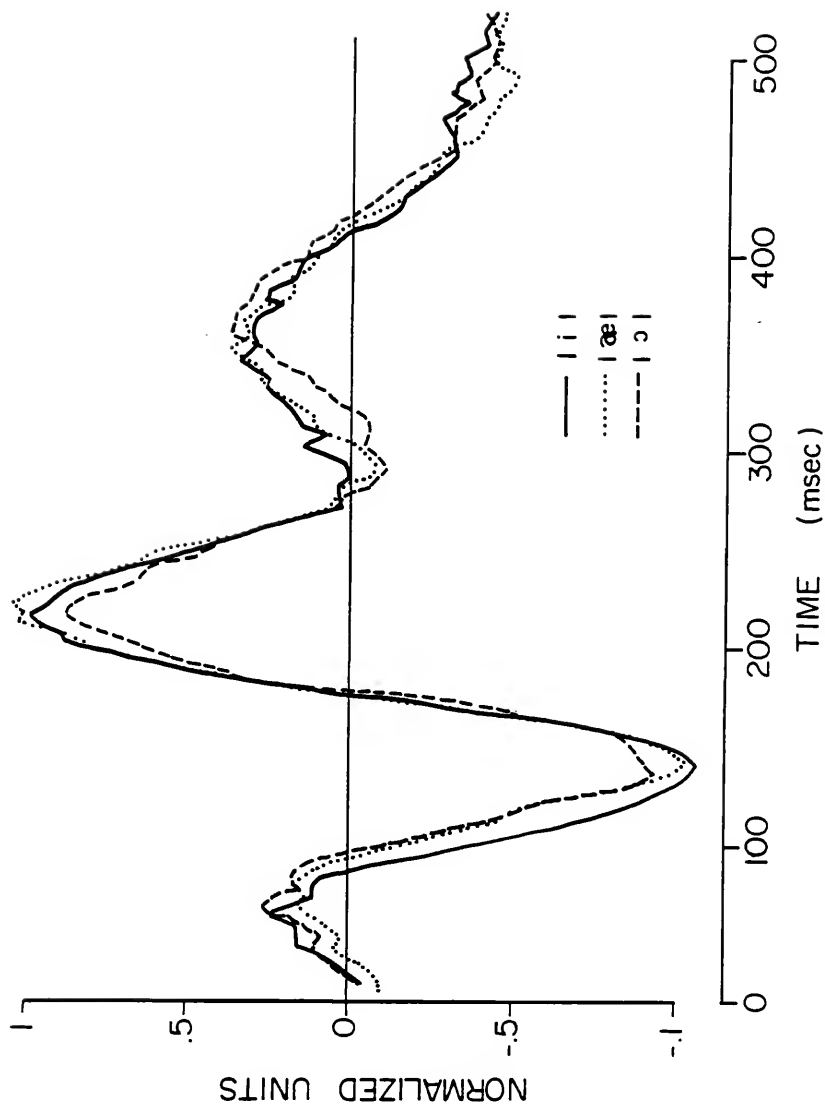


between left and right hemisphere AER's for the entire duration of their time courses (although congruence was observed between 100 to 130 ms, and 265 to 310 ms). The largest differences were noted between 145 and 225 ms, which agrees well with the factor peak latency of 200 ms. In this case too, observed differences in grand mean AER's were accurately represented by factor latencies.

Grand mean AER's for vowels were plotted in Figure 3-7, in order to permit comparison of their differences with factor latencies. Factor 2 (major peak at 75 ms) and Factor 7 (major peak at 115 ms) were shown to differentiate /i/ from /æ/, and /i/ from /ɔ/. Visual inspection of the waveforms plotted in Figure 3-7 revealed that the AER associated with /i/ was most different from the /æ/ and /ɔ/ AER's between 25 and 110 ms. Thus, factor latencies were generally consistent with waveform areas of difference.

Factor 4 (major peak at 200 ms) was shown to differentiate /æ/ from /ɔ/. Figure 3-7 revealed waveform differences between the /æ/ AER and the /ɔ/ AER between 120 to 130 ms, 175 to 215 ms and 285 to 320 ms. Factor 9 (major peak at 290 ms) differentiated between /i/ and /ɔ/, /æ/ and /ɔ/, and /i, æ/ and /ɔ/. Visual inspection of Figure 3-7 revealed that from 175 to 215 ms and from 285 to 320 ms, the /i/ and /æ/ waveforms were most similar to each other, and most different from the /ɔ/ waveform. In both cases, more

Figure 3-7. Grand mean average evoked responses to syllables containing /i/, syllables containing /æ/, and syllables containing /ə/.



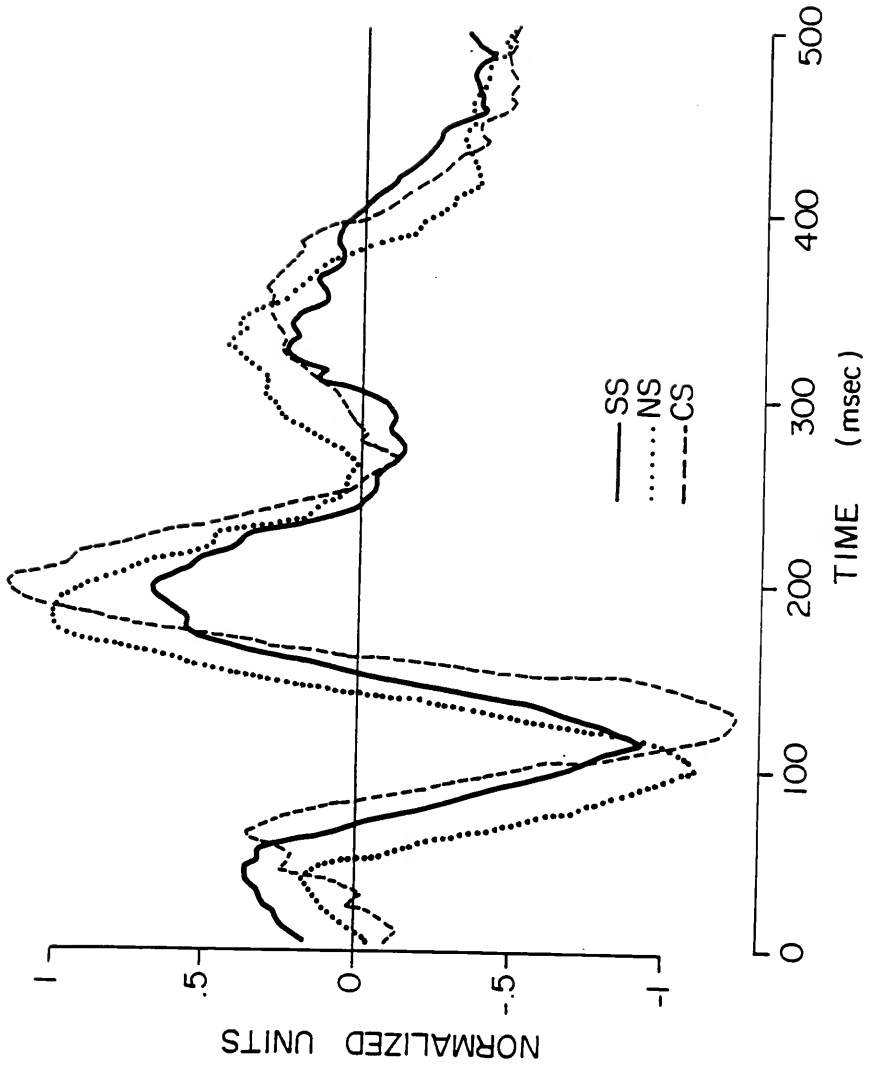
areas of waveform difference were observed between grand mean AER's than were specified by the factor data.

In summary, for the Vowel main effect, observed differences in the AER data between vowels (or pairs of vowels) occurred at many latency epochs. Latencies of factors differentiating between vowels were generally associated with observed differences in grand mean AER's. However, some AER differences were not associated with the significant factor. This discrepancy may have been due to variability in the individual AER data, as described above.

Finally, grand mean AER's for Trial were compared to determine if factor latencies were in the approximate areas of observed differences. Factor 1 (with peaks at 40, 90 and 150) and Factor 2 (with a major peak at 75 ms and a smaller one at 145 ms) differentiated all pairs of trials except CF and CS. Visual inspection of the group averaged AER's for the NS, SS and CS trials (Figure 3-8) revealed large differences between waveforms in the latency area of 40 to 150 ms. (The CF trial was not included in this figure because the CF and CS trials were not found to be significantly different.) Thus, the latencies of Factors 1 and 2 correlated well with actual observed differences in grand mean AER's.

Factor 4 (with a peak at 200 ms) and Factor 10 (with peaks at 375, 445 and 480 ms) differentiated between all pairs of trials except NS and SS (and, of course, between CS

Figure 3-8. Grand mean average evoked responses to synthetic syllables, natural syllables, and chirps with speech instructions.





and CF). Thus, it would be expected that at the latencies associated with Factors 4 and 10, the SS and NS trials would show congruent waveforms. Visual inspection of the group averaged AER's presented in Figure 3-8 revealed that from 220 to 235 ms, 360 to 380 ms and 430 to 485 ms, the SS and NS waveforms showed their greatest similarity. Again, the extracted factors related well to the observed data.

Factor 8 (with peaks at 320 and 420 ms) differentiated between NS and SS trials only. Visual inspection of the data presented in Figure 3-8 revealed that the NS and SS waveforms were markedly different along their entire time courses, with the exception of the latencies associated with Factors 4 and 10. At 320 ms, the CS and SS AER's appeared to be most similar, and separated from the NS waveform; while at 420 ms, the NS and SS waveforms were widely separated, with the CS waveform forming a midline.

In summary, the latencies of observed differences in grand mean AER's based on Consonant, Hemisphere and Trial appeared to be generally consistent with the latencies of their associated factors. Further, latencies of observed similarities in waveforms appeared to be generally consistent with the latencies of factors which did not differentiate them. For vowels, differences in grand mean AER's were observed at the associated factor latencies; however, additional variations in grand mean AER's were also noted at epochs which did not correspond to factor

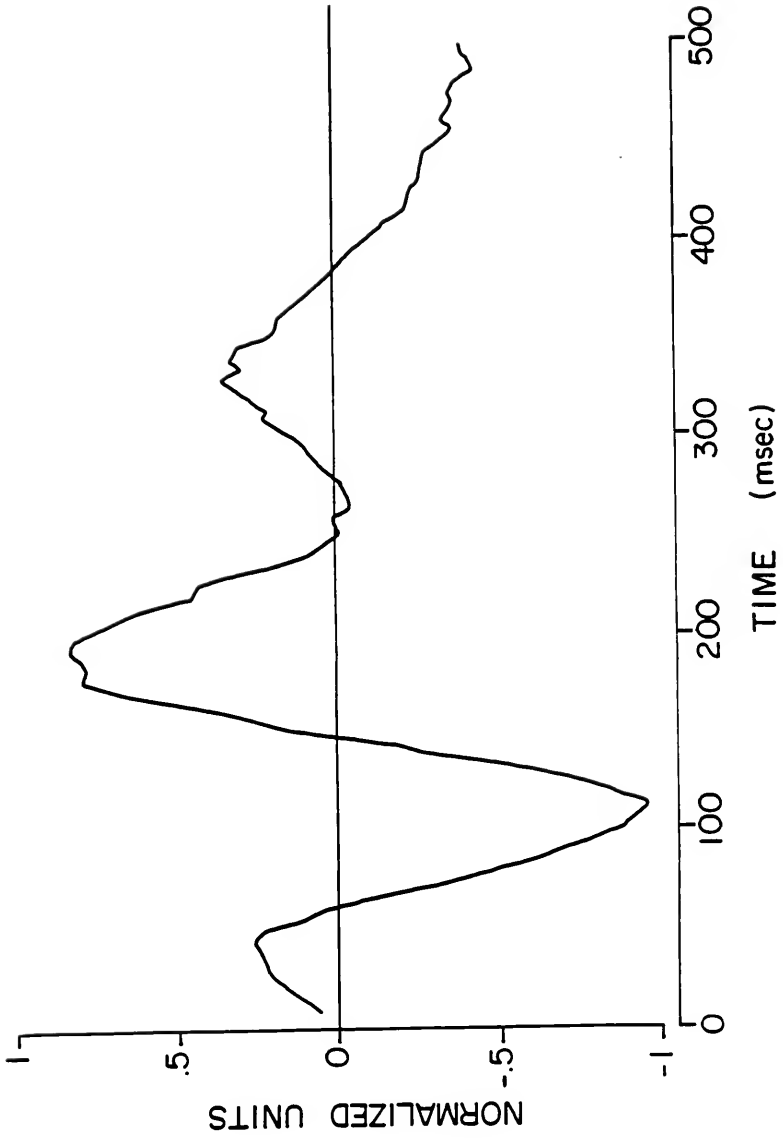
latencies. This discrepancy may have been due to variability in individual AER's. Indeed, such variance might even be expected, given the differences in the "vowel" portion of the syllables and the chirps.

#### Analysis Two: Synthetic and Natural Syllables Only

In order to eliminate the confounding task variables of stimulus difficulty/required attention and incorrect perceptual judgements, the two chirp trials were excluded from the data, and a second PCA was done on the synthetic and natural syllables alone. This design permitted comparison of the PCA and ANOVA results between analyses, so that the effects of the above task variables could be assessed.

In the second analysis, the centroid was based on 288 normalized AER's resulting from 12 subjects x 2 hemispheres x 2 consonants x 3 vowels x 2 trials ( $12 \times 2 \times 2 \times 3 \times 2 = 288$ ). The centroid is pictured in Figure 3-9. It is characterized by a small positive peak at 40 ms (P40), a large negative peak at 115 ms (N115), a large positive peak at 190 ms (P190), a negative peak at 265 ms (N265), a positive peak at 330 ms (P330), followed by a gradual decline asymptoting at 490 ms (N490). This centroid is very similar to the one obtained from the full data set utilized in the first analysis.

Figure 3-9. The centroid or grand mean auditory evoked response obtained from the Principal Components Analysis based on synthetic and natural syllables (Analysis Two).



The PCA was calculated in a manner similar to that described for the full data set. Ten factors were extracted which accounted for 79.7% of the observed variance in the data. The rotated factors are pictured in Figures 3-10. Factor scores were then calculated, and submitted to an Analysis of Variance (ANOVA) program (Dixon, 1981). This procedure resulted in 10 new ANOVA's based on the 10 extracted factors, each evaluating all possible main effects and interactions among the independent variables.

#### Primary Hypothesis Analysis

As in the previous analysis, left hemisphere differentiation of /b/ and /d/ was the issue of primary interest. In Analysis One, left hemisphere /b/ and /d/ mean factor scores were averaged over "speech instruction" trials and compared. In the present analysis, only the syllable trials were included; thus Trial was not a variable of major concern.

Based on the results of Molfese (1980a) and Molfese (1983), it was hypothesized that /b/ and /d/ would be differentiated in the left hemisphere and not the right. In order to test this hypothesis, a Consonant by Hemisphere interaction was sought in the 10 ANOVA's based on the 10 extracted factors. Results revealed that such an interaction significant at the .05 level or better was not present. Thus, when only synthetic and natural syllables

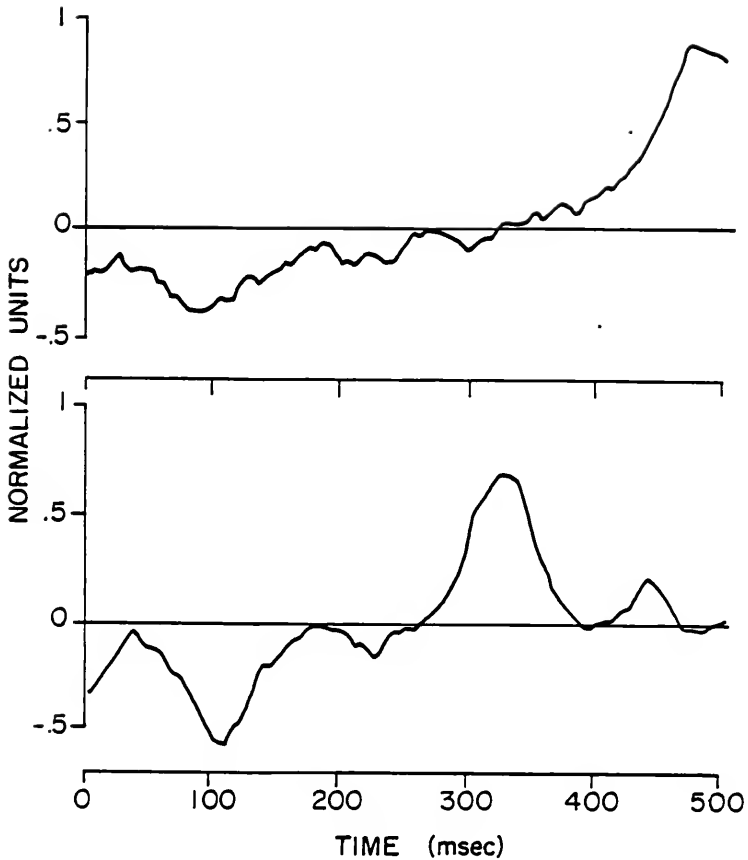
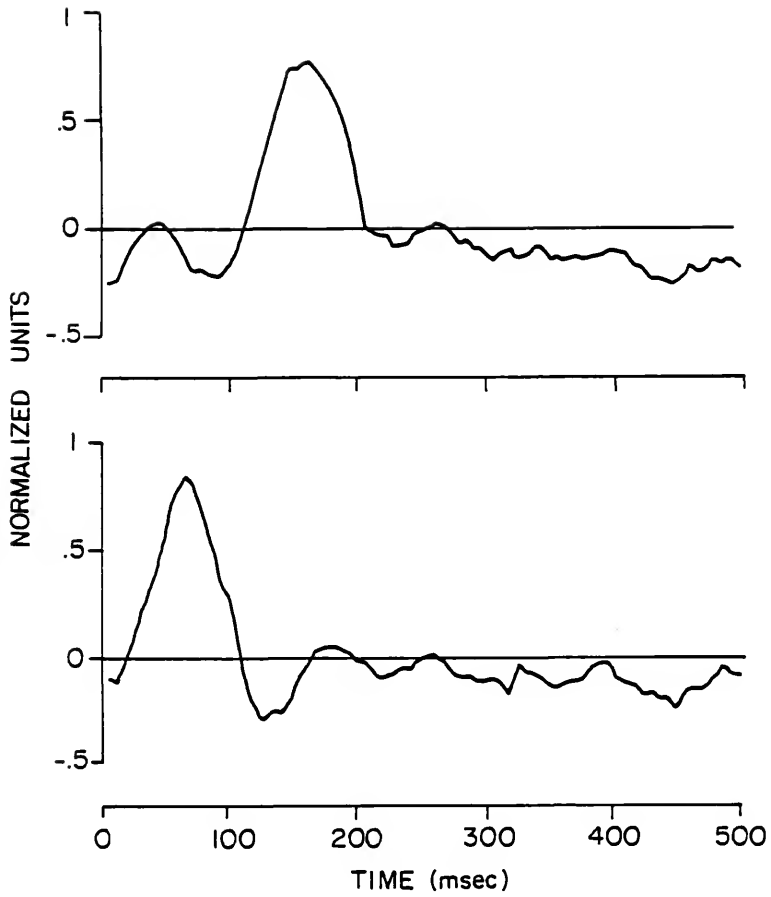
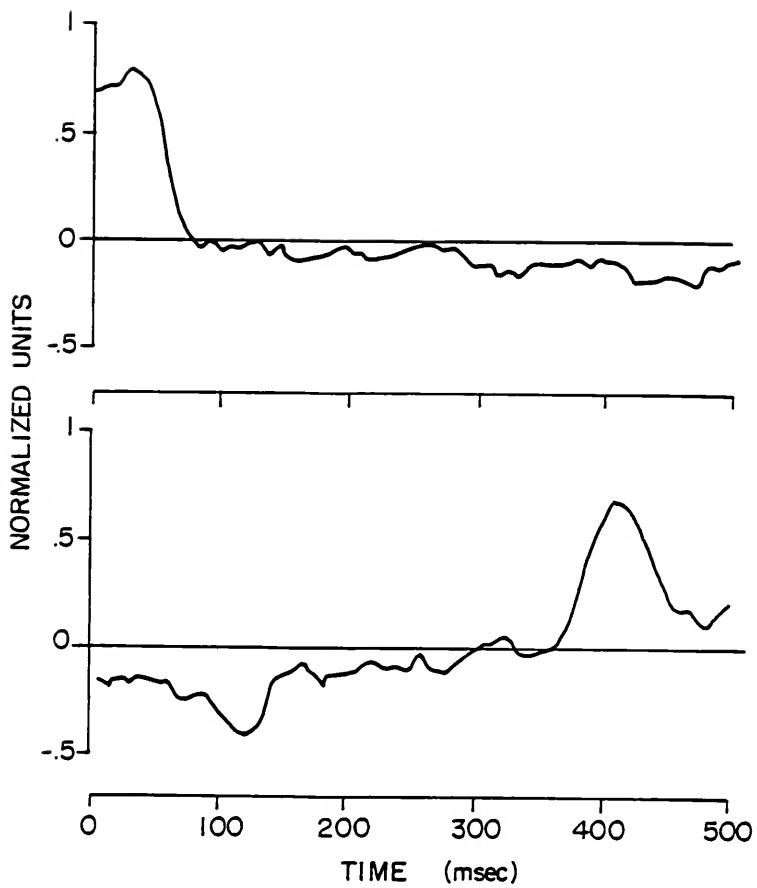
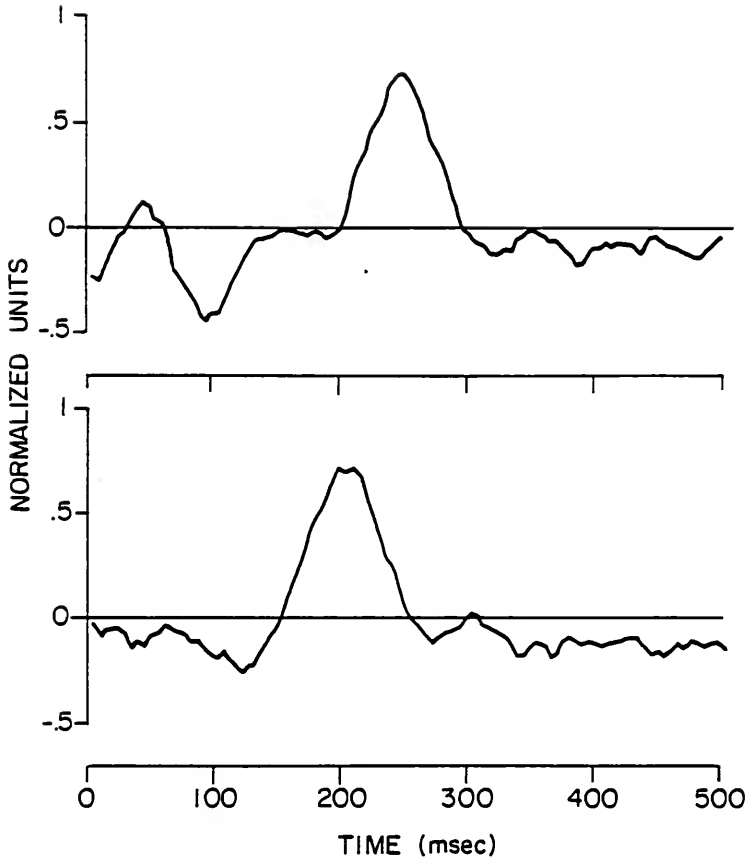


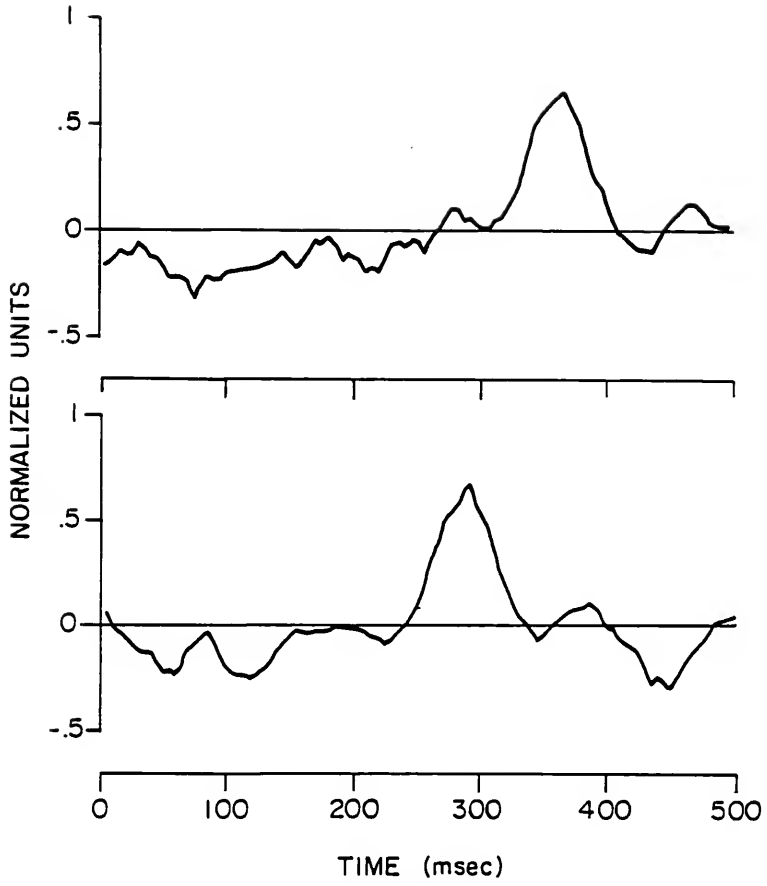
Figure 3-10. The ten factors extracted by means of a Principal Components Analysis based on natural syllables and synthetic syllables. (a) Factor 1, (b) Factor 2, (c) Factor 3, (d) Factor 4, (e) Factor 5, (f) Factor 6, (g) Factor 7, (h) Factor 8, (i) Factor 9, (j) Factor 10











were analyzed, a left hemisphere differentiation between /b/ and /d/ was not obtained.

This finding is not in agreement with previous research, and is somewhat contradictory to the previous analysis. Results of Analysis One revealed a significant Consonant by Hemisphere by Trial interaction, and indicated a tendency (although not significant) toward left hemisphere differentiation of /b/ and /d/ for syllables and speech chirps. When speech chirps were eliminated in Analysis Two, no evidence of hemispheric asymmetry in /b-/d/ discrimination was obtained. Thus, it appeared that the task variables of stimulus difficulty/required attention and/or incorrect perceptual judgements were important in causing asymmetric hemisphere involvement.

#### Secondary Analyses

A number of additional questions were posed in Analysis Two, similar to those investigated in Analysis One. They were as follows:

1) Are there bilateral processes which differentiate /b/ from /d/, regardless of vowel context or trial? Previous research and the results of Analysis One had isolated such a process. In order to address the issue in this instance, a main effect for Consonant was sought. Factor 3 (major peak latency at 160 ms) revealed a main effect for Consonant at the  $p = .0145$  level ( $F = 8.41$ ,  $df = 1,11$ ). Because this term had only two means and they were

significantly different, no further analysis was carried out. This result indicated that /b/ and /d/ were differentiated bilaterally. Further, the latency of this bilateral process (160 ms) was in agreement with the latency determined in the Molfese research (170 ms) and in Analysis One (150 ms). For this effect, elimination of the chirp trials did not appear to alter the obtained pattern of significance and latency.

2) Can responses from the left and right hemispheres be differentiated, independent of other variables? Based on previous research and the results of Analysis One, such a difference was expected. To test this hypothesis, the ANOVA data of Analysis Two was examined for significant Hemisphere main effects. Results revealed two factors (3 and 8) with significant differences between the responses of the left hemisphere and those of the right hemisphere, regardless of consonant, vowel or trial ( $F = 6.07$ ,  $p = .0314$ ;  $F = 8.87$ ,  $p = .0126$ ;  $df$  for both = 1,11). Thus, as expected, the responses of the right hemisphere could be reliably differentiated from those of the left, independent of other variables. Again, elimination of the task variables of stimulus difficulty/required attention and incorrect perceptual judgements did not significantly affect the Hemisphere relationship.

3) Are there bilateral processes which discriminate between vowels, independent of other variables? In order to

answer this question, a main effect for the Vowel variable was sought in the Analysis Two ANOVA's. Results of previous research indicated that such an effect might be found.

Results of this analysis revealed that, as expected, various vowel pairs and groups indeed were differentiated bilaterally. Main effects for Vowel characterized four separate factors--2, 4, 5, and 8 ( $F = 9.06$ ,  $p = .0013$ ;  $F = 3.69$ ,  $p = .0415$ ;  $F = 5.00$ ,  $p = .0162$ ;  $F = 6.60$ ,  $p = .0057$ ;  $df$  for all = 2,22). Post hoc Scheffe analyses of Factor 2 (major latency peaks at 115 ms and 325 ms) revealed significant differences between /i/ vs. /æ/, and /i/ vs. /o/. For Factor 4 (major latency peak at 65 ms), differences between /i/ vs. /æ/ and /i/ vs. /æ, o/ were significant at the .05 level. Factor 5 (major latency peak at 30 ms) showed significant differences similar to Factor 4. Finally, for Factor 8 (major latency peak at 200 ms), /æ/ vs. /o/ and /i, æ/ vs. /o/ were significant. Again, as in Analysis One, significant evidence for bilateral processes which discriminate between vowel pairs and groups was found. Latency results were also similar: in both analyses, the spread vowel /i/ appeared to be discriminated at a relatively early latency (115 ms, 65 ms and 30 ms; although 325 ms was also included). The two close vowels /æ/ and /o/ were differentiated somewhat later during the perceptual process (200 ms). Thus, excluding the chirp

trials did not appear to affect the significance or latency of factors associated with Vowel variables.

4) Are synthetic syllables and natural syllables differentiated bilaterally, independent of other variables? Some evidence for such bilateral differentiation was found in Analysis One at 75, 150 and 420 ms post-stimulus onset. It was hypothesized that differences would be found in the present comparison at similar latencies.

Results revealed that a main effect for Trial characterized Factors 2, 3, 4, and 6 ( $F = 15.71$ ,  $p = .0022$ ;  $F = 15.63$ ,  $p = .0023$ ;  $F = 26.98$ ,  $p = .0003$ ;  $F = 7.53$ ,  $p = .0191$ ;  $df$  for all = 1,11). The latencies associated with these factors were 115 and 325 ms, 160 ms, 65 ms, and 410 ms, respectively. Since each main effect compared only two means, further analysis was judged unnecessary. Thus, it appeared that synthetic and natural syllables were differentiated bilaterally, independent of other variables, as expected based on the previous analysis. In addition, the latencies specified were generally similar to those identified in Analysis One as differentiating between NS and SS trials. Once again, eliminating the chirp trials and associated task variables did not appear to affect the relationships obtained in the Trial main effect.

5) What are the effects of Trial on hemispheric involvement in /b/-/d/ discrimination? Are these consonants differentiated in the left hemisphere for one type of

syllable but not the other? Such a finding would indicate that synthetic and natural syllables do not evoke similar patterns of hemispheric involvement, and a positive relationship for this effect was not expected.

In order to determine the possibility of a difference in hemispheric involvement for consonant discrimination related to syllable type (i.e., synthetic or natural), a significant Trial by Consonant by Hemisphere interaction was sought in the data. Such an interaction characterized Factor 3 and Factor 8 ( $F = 4.84$ ,  $p = .0500$ ;  $F = 6.77$ ,  $p = .0401$ ;  $df$  for both = 1,11). However, when post hoc Scheffe comparisons were made between /b/ and /d/ in the left hemisphere for either type of syllable, no significant differences were found. Neither were significant differences found in the right hemisphere. Thus, it appeared that the pattern of hemispheric involvement in discriminating /b/ from /d/ was not significantly different for synthetic and natural syllables--in both cases, no hemispheric asymmetries in /b/ vs. /d/ discrimination were noted. This issue was not directly addressed in Analysis One, so no comparisons can be made. However, these results do indicate that synthetic syllables are valid alternatives to natural speech, at least for perceptual research involving voiced stop consonants.

6) Are natural syllables and synthetic syllables differentiated unilaterally in the right or left hemisphere?

Results of Analysis One indicated that they were not, however, the effects of removing the chirp data from the analysis were unknown. The purpose of this comparison was to investigate this issue.

To assess hemispheric asymmetry in the differentiation between synthetic and natural syllables, the ANOVA's were scanned for Trial by Hemisphere interactions. Factors 2 and 9 showed such an interaction ( $F = 6.85$ ,  $p = .0240$ ;  $F = 19.51$ ,  $p = .0010$ ;  $df$  for both = 1,11). Although post hoc Scheffe tests did not reveal significant differences between the two trials in either hemisphere for Factor 2, Factor 9 showed significant differentiation between synthetic and natural syllables in the left hemisphere but not the right at the .05 level. In this case, it appeared that natural and synthetic syllables were differentiated unilaterally in the left hemisphere. This finding was not consistent with Analysis One. However, it is possible that the more salient AER differences between the responses to syllables vs. chirps masked the smaller differences between the two syllable trials. Thus, it appeared that in an interaction involving hemispheric asymmetry, the effects of the chirp data and associated task variables did influence the obtained relationships.

7) Is there evidence of hemispheric asymmetry in the perception of vowels? A negative finding was expected for this relationship, based on previous data and the results of



Analysis One. Accordingly a Vowel by Hemisphere interaction was not found in the ANOVA data. Further, post hoc analysis of a Trial by Vowel by Hemisphere interaction which characterized Factor 6 ( $F = 4.32$ ,  $p = .0262$ ,  $df = 2,22$ ) did not reveal significant differences between vowels in either hemisphere for either type of syllable. Thus, the results of this test did not support a hypothesis of hemispheric asymmetry in the perception of (undistorted) vowels. This was consistent with previous research and the findings of Analysis One.

#### Methodological Questions

An informal comparison of the centroids and ANOVA results from Analysis One (containing the full data set), and Analysis Two (containing only the responses to the natural and synthetic syllables) revealed some differences and some similarities. For example, the centroid for the full data set exhibited peak latencies which were 5-10 ms longer than those associated with the syllable data alone; however, the waveshapes were very similar. In terms of the ANOVA results, both Analysis One and Analysis Two revealed similar patterns of significance and generally consistent latencies for the Consonant, Hemisphere, Vowel and Trial main effects. However, relationships which involved hemispheric asymmetry were generally different when results of the two analyses were compared. For example, in Analysis One, a significant Consonant by Hemisphere by Trial

interaction was obtained, and post hoc analysis revealed a tendency for initial consonants in "speech instruction" trials to be discriminated in the left hemisphere (although this relationship was not significant at the .01 level). In Analysis Two, no indication of hemispheric asymmetry in the perception of initial consonants was found. In addition, a significant discrimination between natural and synthetic syllables in the left hemisphere was obtained in Analysis Two, but not in Analysis One. Thus, it appeared that inclusion of unfamiliar, difficult ambiguous stimuli did affect results with respect to latency of centroid peaks and hemispheric asymmetry as revealed in the ANOVA's. Main effects for the independent variables, however, did not appear to be influenced.

### Analysis of Perceptual Data

#### Responses to Synthetic and Natural Syllables

During the electrocortical recording procedure, subjects responded to the synthetic syllables with 97.8% accuracy (range = 93-100%), and to the natural syllables with 98.9% accuracy (range = 96-100%). Of the 32 total errors on the synthetic syllables, 9 occurred on the syllable /bi/, 3 occurred on /bæ/, 3 occurred on /bo/, 9 occurred on /di/, 6 occurred on /dæ/, and 2 occurred on /do/. Of the 16 total errors on the natural syllables, 11

occurred on /bi/, 2 occurred on /b<sub>æ</sub>/, 3 occurred on /b<sub>ɔ</sub>/, and none occurred on /di/, /d<sub>æ</sub>/ or /d<sub>ɔ</sub>/. These results indicated that both sets of syllable stimuli were accurately perceived during electrocortical recording.

#### An Analysis of Response Accuracy in the Chirp Trials

When subjects were instructed to respond to high vs. low onset frequencies (the frequency trial), they responded with 55.3% accuracy (range = 45-73%). When they were instructed to respond to /b/ vs. /d/ (the speech trial), mean accuracy was 57.2% (range = 24-83%). One subject, F-1, responded very differently from the other eleven subjects during the speech chirp trial, obtaining a score of only 24% correct. (However, this subject's score of 59% was near the norm for the frequency chirp trial.) When F-1's perceptual responses were eliminated from the averages, a larger difference between percent correct in the frequency trial and the speech trial was observed--55.0% correct for the frequency trial vs. 60.3% accuracy in the speech trial. These results indicate that the discrimination tasks in both the speech and frequency trials were very difficult.

Since subjects were divided into two groups based on order of instructions--those who heard the frequency instructions first and those who heard the speech instructions first--group averages also were compared. Subjects who heard the frequency trial first averaged 54.8%

accuracy in the frequency trial, and 50.2% accuracy in the speech trial. (When F-1's scores were not included, averages were 53.4% and 55.4% respectively). Subjects who heard the speech trial first averaged 55.8% accuracy in the frequency trial and 64.3% in the speech trial.

Two ANOVA's were performed on these data in order to explore the possibility of significant differences occurring between any of the reported means. The first ANOVA program (Helwig and Council, 1979) utilized the complete data set. Results revealed no significant differences at the .05 level for frequency vs. speech trials ( $F = .17$ ,  $p = .6834$ ,  $df = 1,23$ ), or for a trial by order interaction ( $F = 2.68$ ,  $p = .1172$ ,  $df = 1,23$ ). A second ANOVA, from which F-1's data were eliminated, was also calculated. Results revealed, again, no significant differences between the means of the frequency vs. speech instruction trials ( $F = 1.59$ ,  $p = .2235$ ,  $df = 1,21$ ), or in a trial by order interaction ( $F = 1.44$ ,  $p = .2452$ ,  $df = 1,21$ ). Thus, when only percentage accuracy was taken into account, neither instructions nor the effect of order resulted in significant differences. However, it was hypothesized that the proportions of error for each of the six stimuli might reveal additional information on subjects' perceptual strategies; thus an error pattern analysis was carried out also.

Subjects' Perceptual Response Strategies

Based on an analysis of correct and error response patterns, it was possible to draw some inferences regarding subjects' response strategies. For example, if a subject missed 20 out of 20 /b/'s in the context of the vowel /i/ and 0 out of 20 /d/'s, one might infer that in the context of /i/, this particular subject had a response bias toward /d/. Conversely, the same subject may have missed 0 out of 20 /b/'s and 20 out of 20 /d/'s in the context of the vowel /o/. In this case, one might infer that in the context of /o/, the subject had a response bias toward /b/. Such a response pattern might be directly related to the acoustic properties of the six chirps. For example, isolated F2-F3 transitions (chirps) for /i/ end on the relatively high frequencies characteristic of formants two and three for /i/; thus if a subject consistently responded /d/ in the speech trial or "high" in the frequency trial, it is probable that they were attending to the transition offset frequency as their cue. Isolated F2-F3 transitions (chirps) for /o/ end on the relatively low frequencies characteristic of formants two and three for /o/; thus consistent responses of /b/ or "low" also indicated that the subject used the transition offset frequency as a cue. For /æ/, one would predict less consistent results. Based on the acoustic properties of /æ/, subjects could either respond in

a manner similar to the /i/ context, the /o/ context, or in a random pattern.

In terms of the data gathered in this study, response biases as extreme as those described were not consistently observed. However, inspection of the data indicated that for a number of subjects, response biases based on frequency of transition offset were present, and that these response biases appeared to be more pronounced in the frequency trial. In order to statistically test this observation, two chi square tests were performed (see Table 3-1). The actual errors made by subjects on /b/ vs. /d/ at each level of the vowel (/i/, /æ/, /o/) were compared to error frequencies based on chance level expectations, for both the frequency chirp trial and the speech chirp trial. Results indicated that for chirps with frequency instructions, the observed error pattern was significantly different from chance levels (chi square = 14.350,  $p = .0008$ ,  $df = 2$ ). Inspection of the error frequencies in Table 3-1a shows that in the context of the vowel /i/, subjects had a slight response bias toward /d/, while in the context of /o/, subjects had a slight response bias toward /b/. However, for chirps with speech instructions, the obtained error pattern was not significantly different from expected errors based on chance alone (chi square = 3.279,  $p = .1940$ ,  $df = 2$ ). This relationship is presented in Table 3-1b. Thus, it appeared that for chirps with frequency instructions, subjects

Table 3-1. Observed error response patterns for (a) chirps with frequency instructions and (b) chirps with speech instructions. Each cell contains the frequency of error responses out of a possible 240 trials.

---

(a) Frequency Instructions

Vowel	Consonant			Total Errors
	/b/	/d/		
/i/	129	92		221
/æ/	97	104		201
/ɔ/	96	140		236
total	322	336		658

chi square = 14.35, df = 2, probability = .0008

---

(b) Speech Instructions

Vowel	Consonant			Total
	/b/	/d/		
/i/	107	87		194
/æ/	109	86		195
/ɔ/	114	123		237
total	330	296		626

chi square = 3.28, df = 2, probability = .1940

---

responded on the basis of transition offset frequency rather than on the basis of transition onset frequency, while for chirps with speech instructions, a more random strategy was used.

As previously mentioned, subjects were subdivided into groups based on order of instructions: one group heard the frequency chirps first, followed by the speech chirps, and the other group heard the speech chirps first, followed by the frequency chirps. In order to assess the effect of 1) order of instructions, 2) speech vs. frequency instructions, 3) /b/ vs. /d/, and 4) three vowel contexts on subjects' perceptual strategies, a statistical procedure known as Log-linear modelling was applied. Log-linear models are appropriate when it is desirable to analyze the effects of a number of independent categorical variables on a dependent categorical variable. In such cases, an ANOVA is not appropriate, since the dependent variable in an ANOVA must be interval or ratio. When calculating Log-linear models, the observed frequencies are tabulated as logarithms of the raw data, and thus the expected frequencies can be calculated through the addition and subtraction of terms rather than through multiplication and division (Agresti and Agresti, 1979; Brown, 1981). This method of analysis eventually results in a particular model of dependence-independence relationships which explains the observed data, i.e., the observed frequencies are not



significantly different from the expected frequencies of the model. Results of this procedure revealed that for the group receiving frequency instructions first, a CV,CI model was adequate to explain the obtained pattern of error frequencies. In this model, Consonant and Vowel effects were dependent, and Consonant and Instruction effects were dependent, while Vowel and Instruction effects were independent, controlling for Consonant. The data are presented in Table 3-2a. Based on the obtained frequencies, this model can be interpreted as follows: the Consonant and Vowel variables were dependent because listeners (who heard frequency instructions first) used a transition offset frequency strategy in both instruction trials, thus error frequency for each consonant was dependent on the following vowel. Further, this relationship held true regardless of the instructions given, thus Vowel and Instruction trials were independent, given Consonant. The Consonant and Instruction variables were dependent because listeners made fewer errors on /b/ in the frequency trial, and fewer errors on both consonants in the frequency trial. It appeared that initial attention to frequency cues interfered with developing other perceptual strategies in the speech instruction condition.

Listeners who heard the speech instruction trial first showed a markedly different pattern of error responses. For this group, a CVI model was necessary to explain the

Table 3-2. Frequency of error response for (a) subjects who were first given instructions to discriminate chirps on the basis of frequency of onset ("high" vs. "low") and (b) subjects who were first given instructions to discriminate the chirps as speech (/b/ vs. /d/). Each cell contains the frequency of error response out of a possible 120 trials.

(a) Frequency instructions first				
Instructions	Vowel	Consonant		Total
		/b/	/d/	
speech	/i/	82	42	124
	/æ/	69	43	112
	/o/	53	80	133
total		204	165	369
frequency	/i/	65	44	109
	/æ/	40	59	99
	/o/	46	70	116
total		151	173	324
(b) Speech instructions first				
Instructions	Vowel	Consonant		Total
		/b/	/d/	
speech	/i/	25	45	70
	/æ/	40	43	83
	/o/	61	43	104
total		126	131	257
frequency	/i/	64	48	112
	/æ/	57	45	102
	/o/	50	70	120
total		171	163	334

observed error frequencies. In this model, all classification variables--Consonant, Vowel and Instruction trial--were dependent. Based on the data presented in Table 3-2b, it appeared that subjects did not rely on transition offset frequencies in the speech chirp trial, but did rely on this cue in the frequency chirp trial. In addition, there were more total errors made in the frequency condition. Thus, no variables were independent of the others.

These results indicate that stimulus expectation can significantly influence subjects' perceptual response patterns and strategies. In both the speech chirp trial and the frequency chirp trial, stimuli were identical, yet response patterns changed as a function of instructions to the subjects. Further, the order in which the instructions were given also appeared to significantly influence perceptual strategies. These findings are consistent with those of Schwab (1981) and Nusbaum et al. (1983), and support a theory of stimulus expectation as a determinant of perception.

#### Subjective Impressions

Finally, an informal assessment of listeners' perceptions of their strategies was made. Subjects had been asked at the conclusion of the experimental session if they realized that the two sets of chirp stimuli were the same, and if they used the same strategies in discriminating /b/

from /d/ in the speech condition as they did in discriminating "high" from "low" onsets in the frequency condition. Four of the 12 subjects reported that they never realized the stimuli in the two conditions were the same. Of these four, three felt they used different strategies in the two conditions, and one reported inconsistent strategies. The remaining eight subjects reported that at some time prior to the conclusion of the second chirp condition, they did indeed realize that the two sets of stimuli were the same. Of these eight subjects, three stated that they felt they used different strategies in the two conditions, three reported that they used a similar strategy based on frequency, and two reported inconsistent strategies. When these reports were informally compared to each subjects' individual data, it appeared that the three subjects who reported a similar strategy based on frequency did indeed rely on transition offset frequency in both conditions. However, of the total of six subjects who reported different strategies, four appeared to use a similar strategy in both conditions based on transition offset frequency. These results indicate that questioning subjects about their perceptual response strategies may not be an effective way to determine this variable.

## CHAPTER IV DISCUSSION

The purpose of this study was to evaluate a proposed model of speech perception. Some of the assumptions underlying this model were 1) that stop consonants are differentiated in the left hemisphere, 2) that speech and nonspeech tasks involve different perceptual processes and hemispheric involvement patterns, and 3) that "stimulus expectation" can influence the manner in which ambiguous stimuli are perceived. In addressing the first issue, AER's to the voiced stop consonants /b/ and /d/ in three vowel contexts were analyzed. Both perceptual and average evoked responses to "chirps" (isolated F2-F3 transitions taken from the syllable stimuli) were analyzed in investigating the second and third questions. In general, results did not support consistent left hemisphere involvement in /b/-/d/ discrimination. However, some support for the hypothesis of speech-nonspeech differences in perception and the importance of stimulus expectation was demonstrated.

### Hemispheric Involvement in Stop Consonant Perception

The results of this study did not support a hypothesis of consistent asymmetric hemispheric involvement in stop

consonant perception. The first analysis (which involved the entire data set) revealed only a nonsignificant trend toward /b/-/d/ discrimination in the left hemisphere for syllables and chirps with speech instructions. A second analysis (involving only responses to natural and synthetic syllables) revealed no such trend.

In contrast to expectations, consonants (and vowels) appeared to be primarily differentiated by both hemispheres. While strong hemispheric asymmetry has not generally been associated with vowel perception, a number of researchers had hypothesized left hemisphere dominance in the processing of stop consonants. Liberman et al. (1967) and Liberman and Studdert-Kennedy (1978) proposed that stop consonant perception required a "special speech processor" due to the brief duration and variable acoustic cues associated with these phonemes. The results of the present research indicate that such stimulus characteristics are not sufficient to elicit asymmetric left hemisphere involvement.

In addition to the effects of stimulus parameters on neurological processing, the effects of certain task variables (stimulus difficulty/required attention and incorrect perceptual judgements) also were questioned. This issue was raised because previous research and its replication in the first part of this study involved ambiguous stimuli which were difficult to discriminate, more demanding of the subjects' attention, and often perceived

incorrectly. Results of Analysis One, which included ambiguous stimuli, suggested a trend toward left hemisphere differentiation between /b/ and /d/ (although it was not significant at the .01 level). However, when the responses to ambiguous stimuli were removed in Analysis Two, no indication of hemispheric asymmetry in the perception of /b/ vs. /d/ was obtained.

One possible explanation for these results is that the difficulty of the ambiguous stimuli enhanced hemispheric asymmetry in Analysis One. As discussed earlier, there is electrophysiological evidence that increased stimulus difficulty and required attention increase AER latency (Ritter et al., 1972) and amplitude (Poon et al., 1976; Eason et al., 1969; Hartner and Salmon, 1972). In addition, dichotic listening studies reveal that increased difficulty in discrimination results in greater lateralization of response biases (Weiss and House, 1973; Godfrey, 1974; Kasischke, 1979). Thus, it may be the case that the ambiguous stimuli, being more difficult to discriminate than syllables, resulted in AER's which were somewhat greater in amplitude and/or longer in latency in the left hemisphere. When these AER's to ambiguous stimuli were averaged with AER's from syllable stimuli, it is possible that they skewed the data toward greater left hemisphere differentiation than is the case in normal syllable discrimination.

The results of this study appear to support a hypothesis that stimulus difficulty/required attention may be an important determinant of asymmetric hemispheric involvement. The influence of this task variable might also provide an explanation for the results of Molfese's current research (personal communication), in which /b/ and /g/ syllables were still found to be differentiated in the left hemisphere even when no ambiguous stimuli were included. In that study, /b/ and /d/ were also compared and, as in the present research, no hemispheric asymmetry was noted. These findings suggest that /b/-/d/ discrimination may not be perceptually equivalent to /b/-/g/ discrimination. If the task variable of stimulus difficulty enhances hemispheric asymmetry, as hypothesized above, an explanation for the difference in results between /b/-/g/ and /b/-/d/ may be that the /b/-/g/ discrimination task is more difficult.

The results of Blumstein and Stevens (1980) support the concept that /b/-/g/ is a more difficult discrimination than /b/-/d/. In their exploration of the invariant spectral cues associated with /b, d, g/, they concluded that although /b/ and /d/ could be identified on the basis of the initial 10-20 ms of the phoneme, "a velar tends to be identified with fewer errors if the duration of the stimulus is longer than 10-20 ms, suggesting that a longer time is necessary to build up a representation of a 'compact' onset spectrum in the auditory system" (Blumstein and Stevens, 1980,



pg. 661). The finding that /g/ requires a longer processing period than /b/ or /d/ indicates that /g/ is more difficult to perceive.

The observed latencies of bilateral differentiation of /b/ vs. /d/ and /b/ vs. /g/ also support the idea that /b/-/g/ discrimination is more difficult. In the present study, latencies of 150-160 ms characterized /b/-/d/ discrimination, while according to Molfese (1980a) and Molfese and Schmidt (1983), a latency of 170 ms characterized the /b/-/g/ discrimination. The longer latency for /b/-/g/ discrimination may indicate that it is more difficult than /b/-/d/. In summary, it is possible that a task variable, i.e., the difficulty of the discrimination, is perhaps a more important factor in left hemisphere unilateral processing than the nature of the stimuli, in contrast to the hypotheses of Liberman et al. (1967) and Liberman and Studdert-Kennedy (1978).

Despite the possible importance of task variables in eliciting hemispheric asymmetry, the effect of stimulus variables must also be considered. While task variables may determine whether or not the hemispheres will be differentially engaged, it is possible that the relevant stimulus characteristics may determine which hemisphere is utilized. In a re-analysis and review of dichotic listening studies, Lauter (1983) proposed three stimulus variables which appeared to be important in determining asymmetric

left hemisphere involvement. The first was the bandwidth within which the discrimination takes place. That is, according to Lauter's analysis, discriminations within a relatively narrow bandwidth (550 or 570 Hz) elicited a marked right ear advantage (REA), while discriminations within a broader bandwidth (1060 or 1460 Hz) elicited a small to absent REA. Second, the "dimensions of change" of a stimulus, in terms of stimulus complexity appeared to be a factor. Lauter's results indicated that acoustic stimuli containing simultaneous changes of several parameters (e.g., frequency, duration, intensity) consistently showed a greater degree of REA than stimuli in which only a single parameter was varied. The final variable involved event duration or "rate of change." This term referred to the effective stimulus duration of the salient acoustic cue as opposed to the total duration of the acoustic stimulus. Here, Lauter interpreted her results as supporting the hypothesis that shorter effective durations (or faster rates of change) result in increased REA. Thus, it is possible that task variables--specifically stimulus difficulty/required attention--may be the primary influence in whether or not hemispheric asymmetry will be elicited, however, stimulus variables may determine which hemisphere will be dominant.

Finally, two task variables were originally proposed as possible influences on hemispheric asymmetry: stimulus

difficulty/required attention and incorrect perceptual judgements. While the former has been discussed above in relation to previous research findings, the issue of incorrect perceptual judgements has not received much attention in the literature. Unfortunately, the design of this research did not permit a direct comparison of correct vs. incorrect judgements because both stimulus difficulty/required attention and incorrect judgements covaried. Further investigation is needed to separate the effects of this potentially important task variable.

### Stimulus Expectation

#### Perceptual Results

The results of this study did support the assumption that speech and nonspeech tasks involve different perceptual processes, and that stimulus expectation can influence the manner in which ambiguous stimuli are perceived. However, the importance of stimulus expectation appeared to vary as a function of instruction order: when frequency instructions were presented first, additional speech instructions did not significantly alter subjects' pattern of responses, however, when the speech condition was presented first, later frequency instructions did significantly change subjects' response patterns.

One possible explanation for the effect of instruction order was the brief and nonspeech-like nature of the chirp

stimuli. As discussed previously, these stimuli sounded like clicks, and were not immediately recognizable as speech. It is possible that the chirps were so unlike speech that subjects had difficulty in perceiving them as such, even with specific instructions. In this case, the frequency aspect of the stimuli may have been more perceptually salient than the consonant aspect. Thus, when given frequency instructions first, subjects may have maintained a strategy based on frequency rather than attending to the less identifiable speech dimension. On the other hand, subjects who were instructed to attend initially to the speech information contained in the chirps appeared able to use other cues within the acoustic signal rather than the transition offset frequency in discriminating /b/ from /d/ chirps. This was evidenced by significant differences in error pattern as a function of order of instructions. Thus, it is possible that difficulty of the ambiguous stimuli and its acoustic similarity (or dissimilarity) to speech are both important variables in determining the effects of stimulus expectation on response pattern.

#### Electrophysiological Results

Despite significant differences based on response patterns, the electrophysiological results of this study were somewhat ambiguous. Significant differences in hemispheric asymmetry based on stimulus expectation were not

clearly demonstrated. Such findings do not agree with the results of Wood (1975) and Bartholomeus (1974). Wood reported significant differences in grand mean AER's between /b/ and /g/ in the left hemisphere. However, when subjects were asked to attend to fundamental frequency variations in the same /b-/g/ syllables, left hemisphere differences were not present. Bartholomeus (1974) found similar effects when identical stimuli with different task requirements were utilized in a dichotic listening paradigm. Lateralization biases for melody discrimination, voice discrimination, and discrimination between the names of alphabet letters were assessed, using identical stimuli for all three conditions (strings of alphabet letters sung to various melodies by different singers). Bartholomeus' results revealed a tendency toward left hemisphere superiority for the "verbal" task (discriminating between the letters), no hemispheric asymmetry in discriminating voices and a right hemisphere superiority for discriminating melodies. Again, the importance of task variables in determining hemispheric asymmetry was supported.

The differences in results between these studies and the present research may be due to a number of factors. First, different statistical procedures (Wood, 1975) and research methods (Bartholomeus, 1974) were used in comparison to the present study. In addition, in both the Wood (1975) and Bartholomeus (1974) studies, the "nonspeech"

tasks involved attention to a global attribute present over the entire stimulus duration (fundamental frequency, voice, melody), while the "speech" task involved attention to discrete intervals (epochs) during which acoustic cues for consonant identity were present. In the present study, subjects' attention in the nonspeech trial was directed to a particular epoch of the signal--the initial portion--and subjects were encouraged to ignore the rest of the signal. Presumably, subjects also were attending primarily to the initial portion of each chirp in order to make the consonant discrimination in the speech trial (Blumstein and Stevens, 1980; Stevens and Blumstein, 1978). Thus, in the present study, both the entire acoustic stimulus and the relevant aspect of the acoustic stimulus were identical in the speech and nonspeech chirp trials. In such a task, it is possible that potential differences in hemispheric asymmetry based only on instructions were reduced.

#### Hemispheric Involvement in Vowel Discrimination

Results of this study supported a concept of bilateral cortical processing in the perception of vowels, with no evidence for hemispheric asymmetry. These findings agree well with previous research (Molfese and Schmidt, 1983; Cutting, 1974; Shankweiler and Studdert-Kennedy, 1967). They are also in accord with a hypothesis of stimulus difficulty/required attention as an important determinant of

hemispheric asymmetry. To be specific, vowels are less complex stimuli than stop consonants in terms of frequency content, temporal change and duration. Presumably, they are also easier to discriminate, and thus are processed bilaterally in almost all cases. However, there is evidence from dichotic listening studies that when vowel discrimination is made more difficult by adding noise, a left hemisphere superiority for processing vowels occurs (Weiss and House, 1973). Additional research is needed in order to correlate this trend with electrophysiological evidence. A positive finding would further support the concept of stimulus difficulty/required attention as a major influence on hemispheric asymmetry.

#### Hemispheric Involvement in Stimulus Class Differentiation

In both Analysis One and Two, bilateral processes for discriminating stimulus classes were noted at both early (65 to 160 ms) and late (325 to 420 ms) epochs. Thus, it appeared that the two hemispheres functioned in a similar manner in differentiating acoustically diverse stimuli at various post-stimulus latencies.

Molfese and Schmidt (1983) found that stimulus classes (normal bandwidth syllables vs. sinewave formant analogs) were also differentiated in the early latency range (120 ms) and again somewhat later (270 ms). Although the early latency agrees well with the present study, the second

latency does not. This difference is not surprising, given the diverse nature of the stimuli utilized in the Molfese research and this investigation. It is quite possible that following an early response within the first 150 ms of the stimulus, additional acoustic characteristics, such as bandwidth, duration, the presence of noise bursts, and other variables, are processed by the cortex at different latencies.

In addition to bilateral processes, asymmetric left hemispheric involvement was also found in response to different stimulus classes. Results of Analysis One (the full data set) revealed differential left hemisphere response to syllable stimuli (synthetic and natural) as opposed to ambiguous stimuli (chirps with frequency instructions and with speech instructions). In Analysis Two (synthetic and natural syllables only) differential left hemisphere responses to the natural syllables vs. the synthetic syllables were obtained. Latencies for these left hemisphere processes occurred at 200 ms in Analysis One, and at 40, 185 and 325 ms in Analysis Two. Of course, it is difficult to compare these results to Molfese and Schmidt (1983), because different types of stimuli were used. However, the results of Analysis Two showing left hemisphere differentiation between spoken and synthetic syllables are somewhat similar in terms of latency to Molfese and Schmidt's left hemisphere differentiation between normal and



sinewave formant CV's (they reported 215 and 390 ms). Thus, stimulus classes of similar duration but with different internal structure appear to be differentiated both at latencies of approximately 200 ms (185, 215 ms) and again at approximately 350 ms (325, 390 ms).

The results of this study indicate that classes of stimuli based on duration are differentiated both bilaterally and in the left hemisphere, as are natural vs. synthetic syllables. Patterns of latency suggest that the bilateral processes reflect both stimulus (early) and cognitive (late) variables. The left hemisphere response appears to occur somewhere between these two bilateral processes, and may reflect an intermediate stage of decision-making. Again, this interpretation is highly speculative, and requires further investigation.

The findings discussed in this section relate primarily to the issue of how nonphonetic stimulus differences are processed neurologically. Although specifying such effects may help to clarify the relationship between bilateral and unilateral hemispheric involvement, it does not directly address the speech perception issue.

#### A (Revised) Theory of Speech Perception

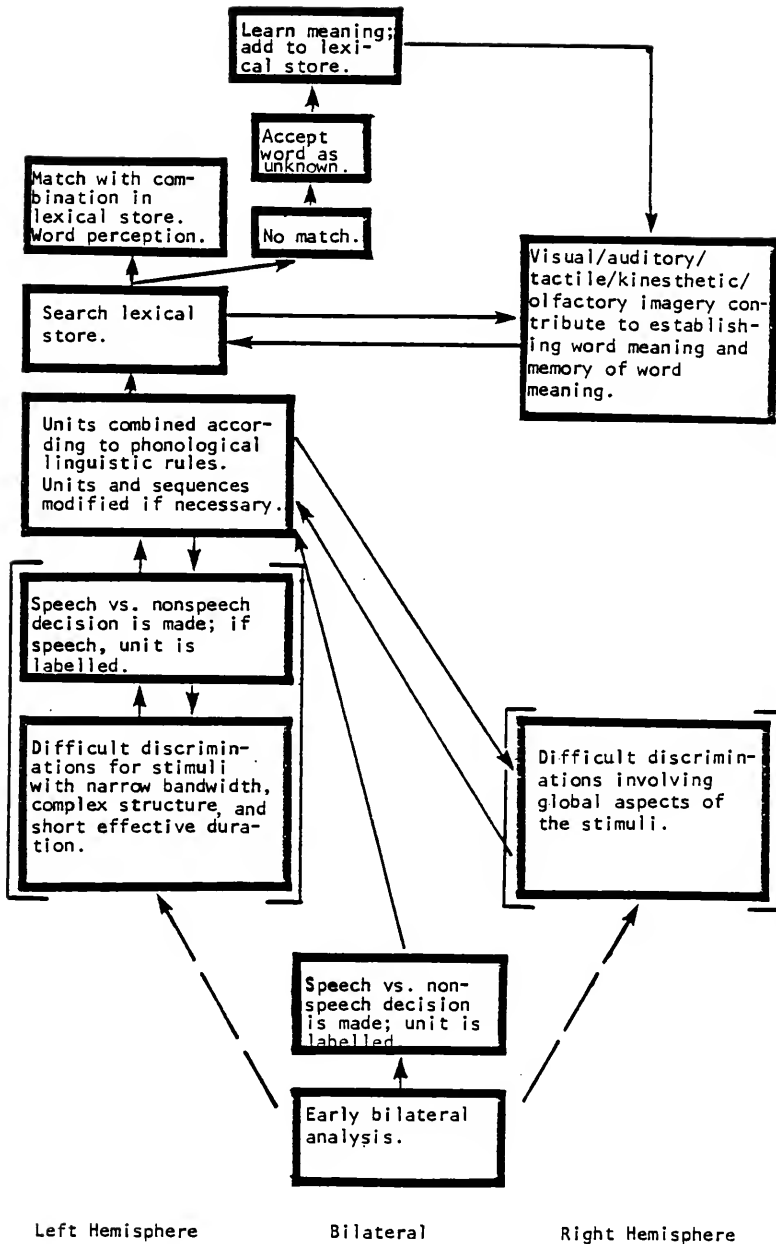
Based on the results of this study, the model of speech perception proposed in Chapter I appears to require modification. In the initial model, it was assumed that the

left and right hemispheres were specialized for different functions during initial acoustic analysis of the signal. The left hemisphere, according to the model, processed "complex, rapidly-changing frequency over time information." In the right hemisphere, longer-duration spectral and temporal analyses were hypothesized to occur. It was further assumed that when a speech signal was anticipated (on the basis of initial acoustic analysis and/or specific instructions, presuppositions, etc.), a series of primarily left hemisphere processes were engaged.

Further, no reference was made in the previous model to task variables, specifically, the difficulty of or attention required for the particular discrimination. The results of this study indicate that these task variables may be important in determining patterns of hemispheric asymmetry. Thus, the previous model of hemispheric involvement in speech perception must be modified to include task variables.

A modified model of speech perception is presented in Figure 4-1. In this model, speech perception at the earliest stages is mediated by bilateral processes. At this stage, it may be the case that the left and right hemispheres are performing identical functions, or it may be that the left and right hemispheres are performing different functions occurring at identical latencies. In either case, both analysis of the linguistic aspect of the stimuli and

Figure 4-1. A (revised) model of speech perception.  
See text for discussion.



its general acoustic characteristics appear to occur bilaterally at early (prior to 170 ms) latencies. In both the present study and the Molfese research (Molfese, 1980a; Molfese and Schmidt, 1983), this bilateral cortical process was observed.

If task variables are such that this early bilateral processing is insufficient to resolve the stimulus as belonging to one group or another, it is proposed that a lateralized process may occur. If the discrimination involves a narrow bandwidth and/or complex stimulus structure and/or rapid rate of change ("effective duration"), as hypothesized by Lauter (1983), the left hemisphere may become involved in stimulus discrimination in a way that the right hemisphere does not. Indeed, the results of Molfese (1980a) and Molfese and Schmidt (1983) lend support to this hypothesis. In both studies, an early bilateral process was observed, but a lateralized left hemisphere differentiation also occurred at a later point in time. The consonantal stimuli, /b/ and /g/, involved complex structure and rapid rates of change (formant transitions), thus conforming to the criteria of acoustic parameters affecting left hemispheric involvement hypothesized by Lauter (1983). However, such a lateralized process was not demonstrated in the present research. Presumably, the /b/ and /d/ stimuli were also complex in structure and contained rapid rates of change, however, it

was hypothesized that they were discriminated more easily than /b/-/g/, and thus involved only the early bilateral cortical process.

Implicit in the assumption of lateralization is the idea that other sets of stimulus characteristics will involve the right hemisphere when bilateral processing is insufficient. This may explain the results of Molfese (1978b) and Molfese (1980b), in which voice onset times (VOT's) at the discrimination boundary were differentiated in the right hemisphere. In order to discriminate VOT, it may be necessary for a listener to attend to the "on-off" characteristics of a portion of the stimulus, rather than the bandwidth, structural complexity or rate of change variables associated (by Lauter, 1983) with left hemisphere involvement. These speculations, however, require further testing and analysis.

As in the original model, "stimulus expectation" is hypothesized to affect perception at these early levels, as evidenced by differences in response patterns based on instructions in both the present study and Schwab (1981). Both the early bilateral analysis and secondary asymmetric hemispheric differentiation can be affected by higher cortical feedback, and this feedback can override actual acoustic cues. If an articulatory referent exists (Liberman et al., 1967), it would be incorporated at this level of perception.

The remainder of the model, which deals primarily with the assignment of meaning to the phonological sequences derived from early analyses, is unchanged. Additional research must be carried out in order to evaluate the significant hypothesized left hemisphere contribution to the semantic process.

In conclusion, it appears that the role of the left hemisphere is not as prominent in consonant discrimination as has been previously hypothesized (Cutting, 1974; Shankweiler and Studdert-Kennedy, 1967), and that asymmetric hemispheric involvement may be a function of task variables as well as stimulus variables. Additionally, the influence of stimulus expectation, based on instructions to the subjects regarding the nature of ambiguous stimuli, should not be overlooked in the analysis of perceptual responses.

#### Conclusions

1) Asymmetric hemispheric involvement results from an interaction of (at least) two variables: stimulus characteristics and task demands. The stimulus characteristics appear to determine which hemisphere will be utilized, while task demands influence whether or not asymmetric processing is necessary.

2) Perception at the phoneme level can be accomplished by means of a bilateral cortical process in most cases.

When the required discriminations are particularly difficult, left hemisphere unilateral processing is engaged.

3) Speech and nonspeech tasks appear to involve different perceptual strategies, and stimulus expectation can influence the manner in which ambiguous stimuli are perceived. However, the stimulus characteristics of the ambiguous stimuli (e.g., their similarity to speech) can determine how effectively different instructions will change perceptual strategy.

4) Electrophysiological differences as a function of stimulus expectation may not be significant if both the stimulus items themselves and the relevant aspect of the stimuli remain constant in the two instruction conditions. Previous research had indicated significant hemispheric differences in the processing of identical stimuli with varying instruction conditions. However, although the acoustic tokens were the same, the subjects' attention was directed to very different aspects of the stimuli in the different trials. When subjects attend to identical aspects of identical acoustic signals with varying instructions, hemispheric differences may not reach significance.



## APPENDIX

### The Origin of the Speech Mode of Perception

If one proposes a model in which speech perception is somehow "special," the question of the origin of this manner of perception must be addressed. For a model such as this, which relies heavily on cognitive factors, the evolution of the human brain and intelligence must be considered in hypothesizing a solution.

It is possible that as humans developed more complex behaviors, such as tool-making, the need for precise communication grew. Presumably, communication at this point facilitated survival of the early humans by allowing them to pool their knowledge and thus deal more effectively with their environment.

In order for speech to have any value at all as a communicative instrument, it first had to be fairly rapid--perhaps not as fast as the speed of thought, but fast enough to allow rapidly-occurring ideas to be exchanged. Thus, it was necessary for the speech perceptual mechanism to be adequate to process short-duration phonemes in a rapidly-changing context. Second, in order for speech to be a useful way of communicating, the perceptual mechanism had to be adequate to process an inevitably degraded signal. Background noise would always be present, to varying degrees. Parts of the acoustic signal would always be

masked. To rapidly perceive speech, a listener would have needed to develop some way of compensating for signal distortion.

It is impossible to say whether communicative need forced the evolution of the human brain, or if evolution of the brain allowed speech and language to develop. Regardless of the answer, the eventual physiology of the human brain permitted compensation for signal distortion in a variety of ways. First, it was possible for a listener to associate the acoustic signal with visual patterns of the speaker's facial movements. Second, the listener was able to associate both the acoustic and visual feedback with the memory of the meanings of those particular patterns. And finally, it may be that as the listener gained more experience as a speaker, it was possible to associate the auditory and kinesthetic patterns the listener himself remembered from his own previous attempts at speech, and of course, the meaning he had sought to communicate.

In this way, meaning became highly associated with auditory, visual, tactile and kinesthetic cues. However, it is possible that the need to derive meaning from the signal was so strong that the incoming acoustic and visual patterns did more than feed upward from the periphery to the cortex. In order to maximize the chances of making the signal meaningful, downward (efferent) pathways from the cortex may have been utilized to modify the incoming stimuli to conform

to previous knowledge of physical and linguistic constraints. Thus, signal distortion was compensated for, and speech became maximally useful.

In the process of developing more complex behaviors, including speech and language, hemispheric asymmetry most likely evolved, with one hemisphere becoming dominant and specialized for certain cognitive functions. Denenberg (1981) reviewed a vast body of research with chicks, songbirds, rats and primates, and concluded that various patterns of hemispheric asymmetry exist in all species studied. Of particular interest was his conclusion that left hemisphere activation "occurs in songbirds and nonhuman primates in response to salient auditory or visual input, or when a communicative output is required" (Denenberg, 1981, pg. 45). Although the research he cites is far from unequivocal, a case can be made for an evolutionary trend toward hemispheric specialization, with the left hemisphere specialized for certain types of auditory perception and speech/language. This line of reasoning is also consistent with the proposed model, which posits extensive left hemisphere involvement both at the acoustic level and in the feedback loops which influence speech perception. At this point, while perceptual research supports the existence of feedback loops and stimulus expectation in speech perception

(Day, 1970; Warren and Warren, 1970; Schwab, 1981; Nusbaum et al., 1983), direct electrophysiological evidence is lacking.

BIBLIOGRAPHY

- Agresti, A. and Agresti, B. (1979) Statistical Methods for the Social Sciences, Dellen, San Francisco.
- Bartholomeus, D. (1974) Effects of Task Requirements on Ear Superiority, Cortex, 10, 215-223.
- Bickerton, D. (1979) Where Presuppositions Come From, in Syntax and Semantics 11, C. Oh and D. Dinneen (Eds.), Academic Press, New York.
- Blechner, M., Day, R., and Cutting, J. (1976) Processing Two Dimensions of Nonspeech Stimuli: The Auditory-Phonetic Distinction Reconsidered, Journal of Experimental Psychology: Human Perception and Performance, 2, 257-266.
- Blumstein, S. and Stevens, K. (1980) Perceptual Invariance and Onset Spectra for Stop Consonants in Different Vowel Environments, Journal of the Acoustical Society of America, 67, 648-662.
- Broadbent, D. (1954) The Role of Auditory Localization in Attention and Memory, Journal of Experimental Psychology, 47, 191-196.
- Brown, M. (1981) Two-Way and Multiway Frequency Tables--Measures of Association and the Log-linear Model, in BMDP Statistical Software, 1981 Edition, W. Dixon (Ed.), University of California Press, Berkeley, California.
- Chapman, R., McCrary, J., Bragdon, H., and Chapman, J. (1979) Latent Components of Event-Related Potentials Functionally Related to Information Processing, in Progress in Clinical Neurophysiology, Volume 6: Cognitive Components in Cerebral Event-Related

Potentials and Selective Attention, J. Desmedt (Ed.),  
Basel-Karger, Bergen, Netherlands.

- Chomsky, N. (1965) Aspects of the Theory of Syntax,  
M.I.T. Press, Cambridge, Massachusetts.
- Chu, C. (1977) Explicit Presupposition as an Underlying  
Representation, in The Third LACUS Forum, R. DiPietro  
and E. Blansitt (Eds.), Hornbeam, New York.
- Cooper, F., Delattre, P., Liberman, A., Borst, J., and  
Gerstman, L. (1952) Some Experiments on the Perception  
of Synthetic Speech Sounds, Journal of the Acoustical  
Society of America, 24, 597-606.
- Cutting, J. (1974) Two Left Hemisphere Mechanisms in Speech  
Perception, Perception and Psychophysics, 16, 601-612.
- Cutting, J. and Rosner, B. (1974) Categories and Boundaries  
in Speech and Music, Perception and Psychophysics, 16,  
564-570.
- Cutting, J., Rosner, B., and Foard, C. (1976) Perceptual  
Categories for Music-like Sounds: Implications for  
Theories of Speech Perception, British Journal of  
Experimental Psychology, 28, 361-378.
- Daniloff, R. and Moll, K. (1968) Coarticulation of  
Lip-rounding, Journal of Speech and Hearing Research,  
11, 707-721.
- Day, R. (1970) Temporal Order Judgements: Are Individuals  
Language-Bound or Stimulus-Bound? Haskins Labs Status  
Reports 21/22, 71-87.
- Day, R. and Wood, C. (1972a) Interactions Between Linguistic  
and Nonlinguistic Processing, Journal of the Acoustical  
Society of America, 51, 79(A).
- Day, R. and Wood, C. (1972b) Mutual Interference Between Two  
Linguistic Dimensions of the Same Stimuli, paper  
presented at the 83rd meeting of the Acoustical Society  
of America, Buffalo, New York.

- Denenberg, V. (1981) Hemispheric Laterality in Animals and the Effects of Early Experience, The Behavioral and Brain Sciences, 4, 1-49.
- Desmedt, J. (1971) The Centrifugal Auditory System, in Symposium on the Neuropsychology of Hearing, 24th International Congress of Physiological Sciences, Munich, 8, E. Heinz, H. Kromphardt, and B. Pfeiffer (Eds.), Springer, Berlin.
- Dixon, W. (1981) BMDP Statistical Software, 1981 Edition, University of California Press, Berkeley, California.
- Donchin, E. and Heffley, E. (1978) Multivariate Analysis of Event-Related Potential Data: A Tutorial Review, in Multidisciplinary Perspectives in Event-Related Brain Potential Research, D. Otto (Ed.), United States Environmental Protection Agency, Washington, D.C.
- Donchin, E., Ritter, W., and McCallum, W. (1978) Cognitive Psychophysiology: The Endogenous Components of the E.R.P., in Brain Event-Related Potentials in Man, E. Callaway, P. Tueting, and S. Koslow (Eds.), Academic Press, New York.
- Eason, R., Hartner, M., and White, C. (1969) Effects of Attention and Arousal on Visually Evoked Cortical Potentials and Reaction Time in Man, Physiology and Behavior, 4, 283-289.
- Gandour, J. and Dardarananda, R. (1982) Voice Onset Time in Aphasia: Thai. I. Perception, Brain and Language, 17, 24-33.
- Gerstman, L. (1957) Perceptual Dimensions for the Fricative Portions of Certain Speech Sounds, doctoral dissertation, New York University.
- Gevins, A., Zeitlin, G., Doyle, J., Yingling, C., Schaeffer, R., Calloway, E., and Yaeger, C. (1979) Electroencephalogram Correlated of Higher Cortical Functions, Science, 203, 665-668.

- Godfrey, P. (1974) Perceptual Difficulty and the Right Ear Advantage for Vowels, Brain and Language, 1, 323-325.
- Goff, W. (1974) Human Average Evoked Potentials: Procedures for Stimulating and Recording, in Methods in Physiological Psychology, Volume 1: Bioelectric Recording Techniques, Part B: Electroencephalography and Human Brain Potentials, R. Thompson and M. Patterson (Eds.), Academic Press, New York.
- Harrison, J. and Howe, M. (1974) Anatomy of the Descending Auditory System (Mammalian), in Handbook of Sensory Physiology, Volume 1, W. Keidel and W. Neff (Eds.), Springer Verlag, New York.
- Hartner, M. and Salmon, L. (1972) Intra-modality Selective Attention and Evoked Cortical Potentials to Randomly Presented Patterns, Electroencephalography and Clinical Neurophysiology, 32, 605-613.
- Helwig, J. and Council, K. (1979) SAS User's Guide, 1979 Edition, SAS Institute, Inc., Cary, North Carolina.
- Howell, P. and Rosen, S. (1983) Production and Perception of Rise Time in the Voiceless Affricate/Fricative Distinction, Journal of the Acoustical Society of America, 73, 976-984.
- Jasper, H. (1958) The Ten-Twenty Electrode System of the International Federation of Societies for Electroencephalography: Appendix to the Report of the Committee on Methods of Clinical Examination in Electroencephalography, Journal of Electroencephalography and Clinical Neurophysiology, 10, 371-375.
- Kaiser, H. (1958) The Varimax Criterion for Analytic Rotation in Factor Analysis, Psychometrika, 23, 187-200.
- Kasischke, K. (1979) An Investigation of Hemispheric Dominance for Simple Auditory Stimuli, doctoral dissertation, Washington University at St. Louis.



- Kewley-Port, D. (1982) Measurement of Formant Transitions in Naturally Produced Stop Consonant-Vowel Syllables, Journal of the Acoustical Society of America, 72, 379-389.
- Kewley-Port, D., Pisoni, D. and Studdert-Kennedy, M. (1983) Perception of Static and Dynamic Acoustic Cues to Place of Articulation in Initial Stop Consonants, Journal of the Acoustical Society of America, 73, 1779-1793.
- Kimura, D. (1961) Cerebral Dominance and the Perception of Verbal Stimuli, Canadian Journal of Psychology, 47, 191-196.
- Klatt, D. (1980) Software for a Cascade/Parallel Formant Synthesizer, Journal of the Acoustical Society of America, 67, 971-995.
- Kuehn, D. and Moll, K. (1972) Perceptual Effects of Forward Coarticulation, Journal of Speech and Hearing Research, 15, 654-664.
- Lane, H. (1965) The Motor Theory of Speech Perception: A Critical Review, Psychological Review, 72, 275-309.
- Lauter, J. (1983) Stimulus Characteristics and Relative Ear Advantage: A New Look at Old Data, Journal of the Acoustical Society of America, 74, 1-17.
- Liberman, A., Cooper, R., Shankweiler, D., and Studdert-Kennedy, M. (1967) Perception of the Speech Code, Psychological Review, 74, 431-461.
- Liberman, A., Harris, K., Hoffman, H., and Griffith, B. (1957) The Discrimination of Speech Sounds Within and Across Phoneme Boundaries, Journal of Experimental Psychology, 54, 358-368.
- Liberman, A. and Studdert-Kennedy, M. (1978) Phonetic Perception, in Handbook of Sensory Psychology: Perception, R. Held and H. Liebowitz (Eds.), Springer Verlag, New York.

- Lisker, L. and Abramson, A. (1964) Across Language Study of Voicing in Initial Stops: Acoustical Measurements, Word, 20, 384-422.
- Mattingly, I., Liberman, A., Syrdal, A., and Halwes, T. (1971) Discrimination in Speech and Nonspeech Modes, Cognitive Psychology, 2, 131-157.
- McGlone, J. and Davidson, W. (1973) The Relationship Between Speech Laterality and Spatial Ability with Special Reference to Sex and Hand Preference, Neuropsychologia, 11, 105-113.
- Miller, J., Wier, C., Pastore, R., Kelly, W., and Dooling, R. (1976) Discrimination and Labelling of Noise-Buzz Sequences with Varying Noise-Lead Times: An Example of Categorical Perception, Journal of the Acoustical Society of America, 60, 586-598.
- Molfese, D. (1978a) Left and Right Hemispheric Involvement in Speech Perception: Electrophysiological Correlates, Perception and Psychophysics, 23, 237-243.
- Molfese, D. (1978b) Neuroelectrical Correlates of Categorical Speech Perception in Adults, Brain and Language, 5, 25-35.
- Molfese, D. (1980a) The Phoneme and the Engram: Electrophysiological Evidence for the Acoustic Invariant in Stop Consonants, Brain and Language, 9, 372-376.
- Molfese, D. (1980b) Hemispheric Specialization for Temporal Information: Implications for the Perception of Voicing Cues During Speech Perception, Brain and Language, 11, 285-299.
- Molfese, D. and Molfese, V. (1979) Hemisphere and Stimulus Differences as Reflected in the Cortical Responses of Newborn Infants to Speech Stimuli, Developmental Psychology, 15, 505-511.

- Molfese, D. and Schmidt, A. (1983) An Auditory Evoked Potential Study of Consonant Perception in Different Vowel Environments, Brain and Language, 18, 57-70.
- Nusbaum, H., Schwab, E., and Sawusch, J. (1983) The Role of "Chirp" Identification in Duplex Perception, Perception and Psychophysics, 33, 323-332.
- Oh, C. and Godden, D. (1979) Presuppositional Grammar, in Semantics and Syntax 11, C. Oh and D. Dinneen (Eds.), Academic Press, New York.
- Ohman, S. (1966) Coarticulation in VCV Utterances: Spectrographic Measurements, Journal of the Acoustical Society of America, 39, 151-158.
- Oldfield, R. (1971) The Assessment and Analysis of Handedness: The Edinburgh Inventory, Neuropsychologia, 9, 97-113.
- Ostreicher, H. and Sharf, D. (1976) Effects of Coarticulation on the Identification of Deleted Consonant and Vowel Sounds, Journal of Phonetics, 4, 285-299.
- Pastore, R., Ahroon, W., Puleo, J., Crimmins, D., Golowner, L., and Berger, R. (1976) Processing Interaction Between Two Dimensions of Nonphonetic Auditory Signals, Journal of Experimental Psychology: Human Perception and Performance, 2, 267-276.
- Penfield, W. and Roberts, L. (1959) Speech and Brain Mechanisms, Princeton University Press, Princeton, New Jersey.
- Peterson, G. and Barney, H. (1952) Control Methods Used in a Study of the Vowels, Journal of the Acoustical Society of America, 24, 175-184.
- Pisoni, D. (1977) Identification and Discrimination of the Relative Onset Time of Two Component Tones: Implications for Voicing Perception in Stops, Journal of the Acoustical Society of America, 61, 1352-1361.

- Poon, L., Thompson, L., and Marsh, G. (1976) Average Evoked Potential Changes as a Function of Processing Complexity, Psychophysiology, 13,43-49.
- Ritter, W., Simson, R., and Vaughn, H. (1972) Association Cortex Potentials and Reaction Time in Auditory Discrimination, Electroencephalography and Clinical Neurophysiology, 33, 547-555.
- Schouten, M. (1980) The Case Against a Speech Mode of Perception, Acta Psychologica, 44, 71-98.
- Schwab, E. (1981) Auditory and Phonetic Processing for Tone Analogs of Speech, doctoral dissertation, State University of New York at Buffalo.
- Shankweiler, D. and Studdert-Kennedy, M. (1967) Identification of Consonants and Vowels Presented to Left and Right Ears, Quarterly Journal of Experimental Psychology, 19, 59-63.
- Shearer, W. (1982) Research Procedures in Speech, Language and Hearing, Williams and Wilkins, Baltimore, Maryland.
- Sidtis, J. (1982) Predicting Brain Organization From Dichotic Listening Performance: Cortical and Subcortical Functional Asymmetries Contribute to Perceptual Asymmetries, Brain and Language, 17, 287-300.
- Stevens, K. and Blumstein, S. (1978) Invariant Cues for Place of Articulation in Stop Consonants, Journal of the Acoustical Society of America, 64, 1358-1368.
- Warren, R. and Warren, R. (1970) Auditory Illusions and Confusions, Scientific American, 223(10), 30-36.
- Weiss, H. and House, A. (1973) Perception of Dichotically Presented Vowels, Journal of the Acoustical Society of America, 53, 51-58.


Wiederhold, M. and Kiang, J. (1970) Effects of Electric Stimulation of the Crossed and Uncrossed Olivocochlear Bundle on Single Auditory Nerve Fibers in the Cat, Journal of the Acoustical Society of America, 48, 966-977.

Wood, C. (1975) Auditory and Phonetic Levels of Processing in Speeh Perception: Neurophysiological and Information-Processing Analysis, Journal of Experimental Psychology: Human Perception and Performance, 104, 3-20.


#### BIOGRAPHICAL SKETCH

Marylou Pausewang Gelfer was born in Bayshore, New York, on May 23, 1953. She grew up in Patchogue, New York. She earned her Bachelor of Science with Honors in 1975 and Master of Arts in 1976 at Northwestern University, Evanston, Illinois, with a major in speech pathology. After working as a speech pathologist for Wassaic Developmental Center in New York and the Marion County Board of Public Instruction in Florida, she entered the University of Florida to pursue a Ph.D. in speech. She is currently working as a post doctoral research associate at the Institute for Advanced Study of the Communication Processes, University of Florida, Gainesville, Florida.


I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Harry F. Hollien, Chairman  
Professor of Speech

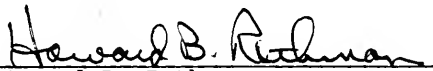
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
W. S. Brown, Jr.  
Professor of Speech

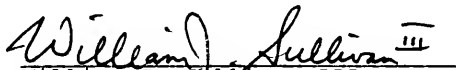
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Ira S. Fischler  
Associate Professor of  
Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
Howard B. Rothman  
Associate Professor of Speech

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
William J. Sullivan III  
Associate Professor of  
Germanic and Slavic Languages  
and Literatures

This dissertation was submitted to the Graduate Faculty of the Department of Speech in the College of Liberal Arts and Sciences and to the Graduate School, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1984

\_\_\_\_\_  
Dean for Graduate Studies  
and Research



UNIVERSITY OF FLORIDA



3 1262 08285 162 6